



Contribution des réseaux de neurones dans le domaine de l'ellipsométrie: Application à la scatterométrie

Issam Gereige

► To cite this version:

Issam Gereige. Contribution des réseaux de neurones dans le domaine de l'ellipsométrie: Application à la scatterométrie. Sciences de l'ingénieur [physics]. Université Jean Monnet - Saint-Etienne, 2008. Français. NNT: . tel-00365631v2

HAL Id: tel-00365631

<https://theses.hal.science/tel-00365631v2>

Submitted on 26 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

en vue de l'obtention du grade de

DOCTEUR de l'Université de Saint-Etienne

École Doctorale : Sciences, Ingénierie, Santé
Spécialité : Optique, Photonique, Hyperfréquences

présentée et soutenue par

Issam GEREIGE

le 5 décembre 2008

Contribution des réseaux de neurones dans le domaine de l'ellipsométrie. Application à la scatterométrie

devant la commission d'examen :

Président :	M. Gérard GRANET	LASMEA Clermont-Ferrand
Rapporteurs :	M. Antonello DE MARTINO	LPICM Paris
	Mme Isabelle SCHANEN	IMEP Grenoble
Examineurs :	Mme Sylvie THIRIA	LOCEAN Paris
	M. Michel STCHAKOVSKY	Horiba JobinYvon
Directeur de thèse :	M. Jean Jacques ROUSSEAU	DIOM Saint-Etienne
Encadrant :	M. Stéphane ROBERT	DIOM Saint-Etienne

Thèse préparée au sein du laboratoire

Dispositifs et Instrumentation Optoélectroniques et Microondes EA 3523

*"Il ne faut pas penser à l'objectif à atteindre, il faut seulement avancer.
C'est ainsi, à force d'avancer, qu'on atteint ou qu'on double
ses objectifs sans même s'en apercevoir."
Bernard Werber, **La révolution des Fourmis***

Remerciements

Qui a dit que la thèse est une empreinte personnelle ? Il est vrai que les tâches accomplies chaque jour au cours de ces trois ans, comme lire un article, tourner une simulation, mesurer un échantillon, etc . . . sont des gestes individuels, mais leur fruit ne peut pas être cueilli mûr sans l'intervention collective d'un certain nombre de personnes spécialisées.

Parmi ces personnes, **Jean-Jacques Rousseau** dont je tiens à remercier pour m'avoir accueilli au sein du laboratoire DIOM (Dispositifs et Instrumentation en Optoélectronique et Microondes) depuis mon stage de Master 2 jusqu'à ma thèse qui l'a dirigée. J'exprime ma reconnaissance envers lui pour ses conseils avisés, sa disponibilité ainsi que son soutien moral et scientifique.

Un remerciement tout particulier que je ne pourrais pas exprimer avec de simples mots, s'adresse à mon encadrant **Stéphane Robert**. Sa rigueur a fait que les travaux présentés au cours de cette thèse ont été menés avec exactitude. Sa disponibilité était un facteur très important qui a permis d'avancer le travail, puisque la porte de son bureau était toujours ouverte pour m'accueillir et répondre à toute question provenant de divers horizons scientifiques et autres. Ses qualités tant scientifiques qu'humaines ont montré son excellence à mener un travail d'encadrement.

Les contributions scientifiques de **Damien Jamon** ont joué un rôle non négligeable dans l'avancement de ce travail. Ses remarques pertinentes m'ont permis d'exploiter au mieux les problèmes rencontrés, grâce à l'originalité de ses idées. Son esprit pédagogique m'a remarquablement aidé pour la structuration du manuscrit ainsi que pour la préparation de la soutenance.

J'adresse également mes sincères remerciements aux membres du jury :

Monsieur **Gérard Granet**, professeur à l'université de Clermont-Ferrand et chercheur au LASMEA, d'avoir présidé ce jury,

Monsieur **Antonello De Martino**, Directeur de Recherche CNRS au laboratoire LPICM de Paris, pour avoir accepté de rapporter sur mes travaux dans les délais imposés par l'école doctorale ainsi que ses remarques fructueuses dans le domaine,

Madame **Isabelle Schanen**, Professeur au Phelma et chercheur à l'IMEP de Grenoble, qui a accepté d'être rapporteur malgré nos exigences en terme de temps,

Madame **Sylvie Thiria**, Professeur à l'université Paris VI et chercheur au laboratoire LOCEAN, qui a accepté d'examiner avec soin ces travaux de recherche et qui a apporté ses conseils utiles dans le domaine des réseaux de neurones,

Monsieur **Michel Stchakovsky**, Ingénieur Horiba Jobin-Yvon, pour avoir accepté d'être membre du jury et examiner ce travail.

Mes remerciements vont également à tous les membres du DIOM qui ont contribué de près ou de loin à ce travail, notamment à **Francois Royer** et **Bernard bayard** pour le temps qui m'ont accordé à relire soigneusement le manuscrit, ainsi qu'à **Bruno Sauviac** pour ses conseils vis-à-vis de la présentation.

Parmi les personnes à qui je dois une reconnaissance particulière, je citerais **Cécile Gourgon**. Je tiens à la remercier d'abord pour la fabrication par nanoimpression ainsi qu'à la caractérisation microscopique des échantillons qui m'ont servi à valider expérimentalement une grande partie de mes résultats. Je la remercie également pour sa patience et sa sympathie durant les 3 premiers mois de mon postdoc, où elle m'a autorisé à revenir régulièrement à Saint Etienne pour finir ma thèse dans les bonnes conditions.

Mes sincères remerciements au Professeur *Fouad Badran* du LOCEAN pour ses conseils et ses vastes connaissances dans le domaine de l'intelligence artificielle.

Je remercie tous les membres du LHC de Saint Etienne avec qui j'ai eu la chance de collaborer, surtout **Stéphanie Reynaud** pour le temps qu'elle m'a accordé pour les mesures AFM.

Que tout l'équipe d'enseignement au sein du département de physique de l'université soit remercié, notamment ceux qui m'ont accompagné dans mon parcours de monitorat. Je citerai **Marie Françoise Blan-Mignon** pour ses services, **Ali Siblani** pour m'avoir suivi de près dans les séances d'électronique, sans oublier **Nicolas Huot** et **Youcef Ouerdanne**.

Je remercie également mes parents pour leur soutien permanent au cours de ces années. Finalement, sans citer des noms, sont profondément remerciés mes amis venant de tous les horizons et qui ont marqué mon passage à Saint Etienne durant ces 4 ans.

Table des matières

Remerciements	1
Introduction générale	19
I Caractérisation ellipsométrique	25
I.1 Ellipsométrie classique	26
I.1.1 Histoire et applications	26
I.1.2 Fondements théoriques	27
I.1.3 Différentes techniques de mesure	31
I.1.3.1 Ellipsomètre à annulation	32
I.1.3.2 Ellipsomètre à modulation par élément tournant	33
I.1.3.3 Ellipsomètre à modulation de phase	34
I.2 Principales structures caractérisables	38
I.2.1 Structure multicouche isotrope	39
I.2.1.1 Présentation et applications	39
I.2.1.2 Formalisme matriciel	39
I.2.2 Structure périodique diffractante	44
I.2.2.1 Présentation et applications	44
I.2.2.2 Phénomène de diffraction par un réseau	45
I.2.2.3 Calcul électromagnétique du champ diffracté	49
I.3 Problème inverse en ellipsométrie	57
I.3.1 Position du problème	57
I.3.2 Formulation mathématique	59
I.3.3 Revue des méthodes d'optimisation	65
I.3.3.1 Méthodes énumératives	65
I.3.3.2 Méthodes déterministes	66
I.3.3.3 Méthodes évolutionnistes	70
I.3.4 Estimation de l'intervalle de confiance	71

I.4	Conclusion	76
II	Réseaux de neurones	79
II.1	Histoire et applications des réseaux de neurones	80
II.2	Principe des réseaux de neurones	82
II.2.1	Du neurone biologique au modèle mathématique	82
II.2.2	Architecture des réseaux de neurones	84
II.2.3	Cas particulier des perceptrons multicouches	86
II.2.3.1	Architecture et définition	86
II.2.3.2	Propriété d'approximation universelle	87
II.2.3.3	Propriété de parcimonie	87
II.2.3.4	Formulation mathématique	88
II.3	Fonctionnement des réseaux de neurones	90
II.3.1	Phase d'apprentissage	90
II.3.1.1	Définition et type d'apprentissage	90
II.3.1.2	Algorithme de rétropropagation de l'erreur	91
II.3.2	Traitement des données d'entraînement	91
II.3.2.1	Initialisation des paramètres internes	92
II.3.2.2	Normalisation des données d'entraînement	92
II.3.2.3	Réduction de la dimension de l'espace des variables	92
II.3.3	Généralisation et Estimation des performances	93
II.3.3.1	Dilemme biais-variance	94
II.3.3.2	Estimation des performances globales du réseau	95
II.3.3.3	Méthodes pour limiter le surajustement	96
II.4	Résolution du problème inverse par réseau de neurones	98
II.4.1	Position du problème	99
II.4.2	Réseau de neurones dédié à la caractérisation ellipsométrique de couches minces de résine	100
II.4.3	Influence de la largeur des domaines d'entraînement	102
II.5	Intervalle de confiance	106
II.5.1	Estimation par réseau de neurones	106
II.5.2	Caractérisation de couches minces de résine	109
II.6	Choix de la signature optique optimale	110
II.6.1	Méthodes classiques d'élagage	111
II.6.2	Méthode de sélection HVS	112
II.6.3	Signature ellipsométrique optimale dans le cas de couches minces de résine	113
II.7	Classification par réseaux de neurones	116
II.7.1	Intérêt de la classification dans la caractérisation optique	116
II.7.2	Le problème de la classification probabiliste	117

II.7.3	La voie neuronale	120
II.7.4	Application à l'identification des gammes d'épaisseurs de couches de résine	123
II.8	Conclusion	126
III	Caractérisation des réseaux de diffraction	129
III.1	Les enjeux de la scatterométrie	130
III.1.1	Besoins des futurs nœuds technologiques	130
III.1.2	Procédés de fabrication	131
III.1.3	Moyens de caractérisation	135
III.1.3.1	Méthodes microscopiques	135
III.1.3.2	Méthodes indirectes	138
III.2	Scatterométrie par réseau de neurones	140
III.2.1	Problème inverse en scatterométrie	141
III.2.2	Etude d'un réseau de silice	143
III.2.2.1	Caractérisations microscopiques	143
III.2.2.2	Caractérisation par réseau de neurones	145
III.2.3	Considérations sur la durée de traitement	148
III.2.4	Caractérisation de différents profils géométriques	149
III.2.4.1	Réseau de profil sinusoïdal	149
III.2.4.2	Réseau de profil trapézoïdal symétrique à bords arrondis	150
III.2.4.3	Réseau de profil trapézoïdal symétrique avec une couche résiduelle	153
III.2.4.4	Réseau de profil trapézoïdal asymétrique	157
III.2.5	Contrôle de l'homogénéité de la structure	160
III.3	Etude paramétrique	162
III.3.1	Paramètres intrinsèques au réseau de neurones	162
III.3.1.1	Influence du nombre de données d'entraînement avec le nombre de neurones cachés	162
III.3.1.2	Domaine de validité du réseau de neurones	163
III.3.2	Paramètres externes au réseau de neurones	164
III.3.2.1	Influence des angles d'incidence θ et d'azimut φ	164
III.3.2.2	Influence de la période	166
III.3.2.3	Influence du choix du modèle géométrique	166
III.4	Sélection de la signature ellipsométrique optimale	173
III.4.1	Estimation des sensibilités de la signature optique	174
III.4.2	Apport de la méthode neuronale	176
III.5	Identification des profils géométriques	180
III.5.1	Etude du réseau en silice	181
III.5.1.1	Validation sur des réseaux simulés	181
III.5.1.2	Résultat expérimental	183

III.5.2	Détection d'une couche résiduelle au fond du réseau	184
III.5.2.1	Validation théorique	184
III.5.2.2	Résultat expérimental	186
III.6	Conclusion	187
Conclusion et perspectives _____		191
Annexe A : Indices optiques des différents matériaux utilisés _		195
Annexe B : Algorithme de rétropropagation de l'erreur _____		199
Bibliographie _____		202

Table des figures

I.1	<i>Nombre de publications issues des précédentes conférences sur l'ellipsométrie, concernant deux catégories : les applications et l'instrumentation (ref : D. E. Aspnes, Thin Solid Films 455-456 (2004)).</i>	27
I.2	<i>Réflexion et transmission d'une onde plane incidente, avec un angle θ_0, sur une surface plane séparant deux milieux 0 et 1 d'indice respectif N_0 et N_1. On montre sur la figure les composantes s et p du champ électrique des ondes incidente (E_i), réfléchie (E_r) et transmise (E_t) ainsi que les vecteurs d'onde associés (\vec{h}_i, \vec{h}_r et \vec{h}_t).</i>	28
I.3	<i>Angles ellipsométriques ψ et Δ en fonction de l'angle d'incidence θ_0 dans le cas des interfaces air-silice et air-silicium pour une longueur d'onde $\lambda = 350$ nm.</i>	30
I.4	<i>Montage ellipsométrique de base (PCSA) comportant un polariseur, un compensateur, l'échantillon et l'analyseur. Les pointillés représentent les axes de transmission de chaque composant optique orienté par rapport au plan d'incidence (représenté par le plan formé par le faisceau lumineux et la normale à l'échantillon) suivant les angles θ_P, θ_C et θ_A.</i>	31
I.5	<i>Représentation schématique de l'ellipsomètre spectroscopique à modulation de phase (UVISEL de Jobin Yvon).</i>	35
I.6	<i>Propagation d'une onde plane à travers une structure multicouche constituée de m couches d'épaisseurs respectifs d_m. θ_0 est l'angle d'incidence dans le milieu 0; θ_i et θ_{m+1} sont les angles de réfraction respectivement dans la couche i et dans le substrat. On note E^+ et E^- les amplitudes des ondes se propageant respectivement dans le sens positif (allant du milieu 0 vers le substrat) et négatif. On désigne par z_i l'abscisse de chaque interface séparant le milieu i et $i + 1$.</i>	40
I.7	<i>Réflexion et transmission d'une onde plane à travers une couche d'épaisseur d et d'indice N_1 déposée sur un substrat d'indice N_2. L'angle d'incidence sur l'interface du film est noté θ_0. Les angles de réfraction dans le film et le substrat sont notés θ_1 et θ_2.</i>	42

I.8	<i>Intensités I_s (a) et I_c (b) calculées dans la gamme du visible au proche infrarouge, pour deux échantillons théoriques constitués respectivement d'une couche de résine d'épaisseur 200 nm et 300 nm, déposée sur un substrat de silicium.</i>	43
I.9	<i>Modèles de réseau de diffraction lamellaire(a) 1D, (b) 2D.</i>	44
I.10	<i>Représentation du phénomène de diffraction par un réseau de période Λ dans l'espace direct en configuration classique.</i>	46
I.11	<i>Représentation dans l'espace réciproque de la diffraction par un réseau de période Λ.</i>	47
I.12	<i>Représentation du phénomène de diffraction par un réseau de période Λ dans l'espace direct en configuration conique.</i>	47
I.13	<i>Découpage du profil géométrique quelconque d'un réseau en couches lamellaires.</i>	51
I.14	<i>Schéma classique d'un réseau de diffraction de profil trapézoïdal symétrique défini par 3 paramètres : la projection de la pente b_1, la largeur de la ligne b_2 et la hauteur h. La période Λ est supposée connue.</i>	52
I.15	<i>Convergence (C_M) de la MMMFE en fonction de l'ordre de troncature M, étudiée dans le cadre d'une signature ellipsométrique (I_s, I_c) sur deux réseaux de profil trapézoïdal pour différentes valeurs de la période Λ. R_a et $R_{l/e}$ sont maintenus fixe à respectivement 0,43 et 0,36. Le nombre de couches est fixé à 20.</i>	54
I.16	<i>Variation de $E_{(I_s, I_c)}$ en fonction de M dans le cas du réseau de profil trapézoïdal en silicium ($b_1 = 30$ nm, $b_2 = 200$ nm, $h = 260$ nm) de période 1000 nm.</i>	54
I.17	<i>Signature ellipsométrique (I_s, I_c) calculée pour un réseau de profil trapézoïdal en silicium ($b_1 = 30$ nm, $b_2 = 200$ nm, $h = 260$ nm) de période 1000 nm, pour différents ordre de troncature.</i>	55
I.18	<i>Représentation de deux profils de période $\Lambda = 565$ nm : trapézoïdal (P_1) et trapézoïdal présentant des arrondis en haut et en bas (P_2).</i>	55
I.19	<i>Convergence C_N de la méthode en fonction du nombre de couches N, étudiée dans le cadre d'une signature ellipsométrique (I_s, I_c) sur un réseau de profil trapézoïdal pour différentes valeurs de périodes.</i>	56
I.20	<i>Evaluation du temps (CPU time) nécessaire pour le calcul d'une signature optique (I_s, I_c) pour (350 nm $< \lambda < 700$ nm) par pas de 10 nm, d'un réseau de diffraction en silice de profil trapézoïdal ($b_1 = 30$ nm, $b_2 = 120$ nm, $h = 180$ nm) et de période 565 nm.</i>	57
I.21	<i>Allure schématique globale d'une fonction objectif montrant l'existence de minima locaux (fort et faible) et d'un minimum global.</i>	61
I.22	<i>Valeur du χ^2 en fonction de l'épaisseur d pour différentes signatures optiques à λ variable paramétrées par θ_0 dans le cas (a) d'une couche de SiO_2 d'épaisseur $d = 500$ nm déposée sur un substrat de Si (b) d'une couche de résine de la même épaisseur déposée sur un substrat de BK_7. (c) et (d) représentent l'erreur χ^2 extraite des figures (a) et (b) pour quelques valeurs significatives de l'angle d'incidence.</i>	62

I.23	Valeur du χ^2 en fonction de l'épaisseur d pour différentes signatures optiques à θ_0 variable paramétrées par λ dans le cas (a) d'une couche de SiO_2 d'épaisseur $d = 500$ nm déposée sur un substrat de Si (b) d'une couche de résine de la même épaisseur déposée sur un substrat de BK_7 . (c) et (d) représentent l'erreur χ^2 extraite des figures (a) et (b) pour quelques valeurs de la longueur d'onde.	64
I.24	Valeurs de χ^2 , dans le cas d'une couche de résine d'épaisseur $d = 500$ nm déposée sur un substrat de BK_7 , en fonction de l'épaisseur de la couche lorsque la signature est bruitée et composée des intensités I_s et I_c pour $\lambda = 400$ nm et $45^\circ < \theta_0 < 80^\circ$ par pas de 5° pour différentes valeurs de σ	65
I.25	Distributions de student à ν degrés de liberté, représentation de la distribution normale ($\nu = \infty$)	73
I.26	Ecart type des intensités I_s et I_c estimé à partir de 10 mesures expérimentales effectuées sur deux échantillons constitués d'une couche de résine d'épaisseur (a) 350 nm noté E_{R_1} et (b) 800 nm noté E_{R_2} , déposée sur un substrat de verre BK_7	74
I.27	Valeurs moyennes des intensités I_s et I_c calculés à partir de 10 mesures expérimentales effectuées sur deux échantillons constitués d'une couche de résine d'épaisseur (a) 350 nm E_{R_1} et (b) 800 nm E_{R_2} , déposée sur un substrat de verre BK_7	75
I.28	Ajustement des intensités expérimentales I_s et I_c par l'algorithme de Levenberg-Marquardt, dans les deux cas où σ est prise constante et celui où σ est mesurée expérimentalement.	75
II.1	Modélisation structurale des échantillons composés d'une couche de résine d'indice N et d'épaisseur h , déposée sur un substrat d'indice connu.	80
II.2	Représentation schématique d'un neurone biologique.	82
II.3	Représentation d'un neurone mathématique. Les variables x_i désignent les entrées, la sortie est notée s_j , le biais b_j et les poids synaptiques sont désignés par $w_{j,i}$	83
II.4	Exemples de fonctions de transfert g : (a) seuil, (b) linéaire, (c) sigmoïde (logistique $K=1$) et (d) Gaussienne.	85
II.5	Architecture d'un réseau de neurones non bouclé complètement connecté à n entrées $N_s - 1$ neurones cachés et un neurone de sortie. w désigne les connections entre les entrées et tous les neurones et z ceux entre les neurones.	86
II.6	Structure d'un réseau de neurones de type PMC à deux couches cachées.	87
II.7	Sortie s calculée par un (a) PMC à deux entrées x_1 et x_2 variant dans l'intervalle $[-4; 4]$, à une couche cachée (fonction d'activation sigmoïdale) et à une sortie (fonction d'activation linéaire). Les figures (b) et (c) correspondent respectivement à s pour $N_c = 2$ et $N_c = 15$, lorsque les poids sont fixés.	89
II.8	Représentation schématique du principe de l'ACP : Réduction de l'espace initial de dimension N en un espace de dimension k tel que ($k < N$), en conservant toutes les données initiales.	93

II.9	<i>Approximation d'une fonction $R(x)$ à partir des données bruitées de celle ci (points) (a) par un modèle linéaire simple $F(x)$, (b) par un modèle polynomial plus complexe $G(x)$.</i>	94
II.10	<i>Principe de la validation croisée; les parties grises sont utilisées pour la validation après chaque apprentissage.</i>	96
II.11	<i>Evolution schématique des performances calculées sur les bases des données d'apprentissage E_{entr} et de validation E_{val}.</i>	97
II.12	<i>Représentation schématique de la caractérisation des propriétés optiques et géométriques de couches minces par réseau de neurones.</i>	99
II.13	<i>Représentation des lois de dispersion de 300 échantillons dont les paramètres sont compris dans les domaines de variation : $1,57 < A < 1,63$; $0,2 \text{ nm}^2 < B < 1,2 \text{ nm}^2$; $0,4 \text{ nm}^4 < C < 1,6 \text{ nm}^4$.</i>	101
II.14	<i>Evolution de l'erreur quadratique moyenne pondérée sur les paramètres au cours de l'apprentissage évaluée sur les corpus d'entraînement, de validation et de test jusqu'à l'arrêt de l'algorithme d'apprentissage par la méthode d'arrêt prématuré.</i>	101
II.15	<i>Représentation des différents paramètres opto-géométriques A, B, C et h calculés par le PMC à partir des signatures bruitées de 200 échantillons test en fonction de leurs paramètres exacts connus.</i>	103
II.16	<i>Erreur quadratique moyenne E_{test}^n sur l'ensemble des échantillons test.</i>	104
II.17	<i>Variation relative des performances théoriques globales des paramètres A, B, C et h lorsque le PMC est entraîné pour différents domaines de h.</i>	104
II.18	<i>Représentation schématique de l'espace engendré par 3 variables (A, B et C) dans un système d'axes orthonormés.</i>	105
II.19	<i>Représentation schématique du processus d'entraînement de la variance.</i>	107
II.20	<i>Résultats concernant l'apprentissage pour un RN de données bruitées (o) issues d'une fonction déterministe (en pointillé). Les sorties du RN_1 sont représentées (en traits pleins) agrémenté des sorties du RN_2 estimant les barres d'incertitude à 1σ.</i>	108
II.21	<i>Loi de dispersion calculée par le RN pour l'échantillon simulé ($A=1.585$, $B=0.65 \text{ nm}^2$, $C=0.85 \text{ nm}^4$). Les barres d'incertitudes sont données à 1σ près.</i>	109
II.22	<i>Evolution de l'erreur quadratique moyenne en fonction du nombre des (a) entrées N du nombre de (b) neurones cachés N_c.</i>	114
II.23	<i>Schéma illustrant le phénomène de classification d'une signature ellipsométrique selon K classes distinctes.</i>	117
II.24	<i>Interprétation géométrique de la règle de décision de Bayes; la partie grise représente la probabilité de commettre un mauvais classement.</i>	118
II.25	<i>Densité de probabilité des épaisseurs pour les échantillons appartenant aux deux classes A et B.</i>	119
II.26	<i>Probabilité a posteriori d'appartenance aux classes C_1 et C_2 en fonction de l'épaisseur h des échantillons.</i>	120
II.27	<i>Probabilités a posteriori calculées par le réseau de neurones (traits+cercles) comparées à celles données par la règle de Bayes (traits continus) pour différents nombres de neurones et d'exemples d'entraînement.</i>	122

II.28	<i>Schéma illustrant le principe de la classification neuronale appliquée à l'identification du domaine d'épaisseur d'une couche de résine déposée sur un substrat.</i>	123
II.29	<i>Sorties du réseau de neurones classifieur (s_1, s_2 et s_3), représentant les probabilités a posteriori d'appartenance aux différentes classes.</i>	124
II.30	<i>Valeurs de l'épaisseur h calculées par optimisation classique (algorithme de Levenberg-Marquardt) en fonction de l'épaisseur initiale h_0 (point de départ) dans le cas de l'échantillon E_1.</i>	126
III.1	<i>Evolution de la fabrication selon la loi de Moore connue en microélectronique selon les prévisions de l'ITRS (International Technology Roadmaps for Semiconductor) en 2004.</i>	131
III.2	<i>Etapes de gravure d'un matériau à travers un masque de résine.</i>	131
III.3	<i>Principe de la lithographie optique par projection.</i>	132
III.4	<i>Principe de la lithographie électronique : écriture directe par un faisceau d'électrons sur une couche de résine électro-sensible.</i>	133
III.5	<i>Principe de la lithographie par nanoimpression thermique.</i>	134
III.6	<i>Influence de la gravure sèche anisotrope et humide isotrope sur la forme du profil final.</i>	135
III.7	<i>Image MEB d'un réseau en résine déposé sur un substrat d'or (Source : Laboratoire Hubert Curien, Saint-Etienne).</i>	136
III.8	<i>Principe de fonctionnement du microscope à force atomique.</i>	137
III.9	<i>Influence de la forme de la pointe AFM sur la mesure du profil géométrique. Les traits pointillés et continus désignent respectivement le profil mesuré et le profil réel.</i>	138
III.10	<i>Principe de la scatterométrie : à partir de la signature optique mesurée sur une structure périodique, les dimensions du profil géométrique sont caractérisées par résolution d'un problème inverse.</i>	139
III.11	<i>Solutions potentielles envisagées pour les futures nœuds technologiques, d'après les prévisions de l'ITRS.</i>	139
III.12	<i>Estimation des paramètres géométriques (a) b_1, (b) b_2 et (c) h par le PMC, ainsi que les barres d'incertitude représentées à 1σ, pour un ensemble de 20 réseaux simulés compris dans le sous domaine test du corpus.</i>	142
III.13	<i>Images du réseau en silice provenant de deux microscopes électronique à balayage différents notés (a) MEB1 (Department of Physics, Université de Joensuu, Finlande) et (b) MEB2 (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles).</i>	144
III.14	<i>Images en 3 dimensions mesurées par deux microscopes à force atomique différents notés (a) AFM1 (Laboratoire Hubert Curien, Saint Etienne) et (b) AFM2 (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles), ainsi que l'extraction sur quelques périodes des profils géométriques à partir de chacune de ces images.</i>	144
III.15	<i>Détermination respective des paramètres géométriques du réseau de diffraction (b_1, b_2 et h) ainsi que sa période Λ par 2 méthodes neuronales (a) utilisant un seul PMC, noté RN_2 (b) utilisant 2 PMC, notés RN_{31} et RN_{32}.</i>	146

III.16	Représentation de l'erreur quadratique moyenne calculée sur I_s et I_c à partir de 92 mesures non utilisées pour les profils reconstruits à partir de la méthode neuronale (RN_1 , RN_2 et RN_3), l'AFM (AFM 1 et AFM 2) et le MEB (MEB 1 et MEB 2).	147
III.17	Temps d'entraînement du PMC en fonction du nombre n d'exemples d'apprentissage et du nombre N_c de neurones.	148
III.18	Profil sinusoïdal de période Λ défini par deux paramètres géométriques b et h .	149
III.19	(a) Résultat d'une caractérisation AFM du réseau sinusoïdal sur une ligne de $10\ \mu\text{m}$, ainsi qu'une comparaison (b) sur une période, entre le profil AFM et celui estimé par le PMC.	150
III.20	Profil trapézoïdal à bords arrondis défini par 5 paramètres : la projection de la pente b_1 , la largeur de la ligne b_2 , la hauteur h et les rayons de courbures bas et haut (respectivement R_1 et R_2).	151
III.21	(a) Profils géométriques obtenus par RN_a et RN_b , (b) Erreur quadratique moyenne calculée pour différentes configurations autre que celle utilisé pour l'entraînement des RN.	152
III.22	Profil trapézoïdal avec une couche résiduelle au fond des sillons, défini par 4 paramètres : b_1 , b_2 , h et l'épaisseur de la couche résiduelle h_r sur le substrat.	153
III.23	Images de profil obtenues par une caractérisation MEB sur le réseau en résine noté $E_{h_r}^1$.	154
III.24	Paramètres géométriques calculés pour différentes valeurs de l'angle d'azimut φ par 4 caractérisations neuronales distinctes.	155
III.25	Déplacement de la matière lors du pressage dans un processus nanoimpression.	156
III.26	Profil trapézoïdal asymétrique défini par 4 paramètres géométriques : les projections des pentes gauches et droites, notées respectivement b_{1g} et b_{1d} , la largeur de ligne b_2 et la hauteur h .	157
III.27	(a) Représentation de 3 profils géométriques asymétriques particuliers : A_1 ($b_{1g} = 10\ \text{nm}$, $b_{1d} = 50\ \text{nm}$), A_2 ($b_{1g} = 30\ \text{nm}$, $b_{1d} = 30\ \text{nm}$) et A_3 ($b_{1g} = 50\ \text{nm}$, $b_{1d} = 10\ \text{nm}$), la hauteur h et la largeur de la ligne b_2 sont égales respectivement à $180\ \text{nm}$ et $120\ \text{nm}$, la période étant $\Lambda = 565\ \text{nm}$; (b) les signatures ellipsométriques I_s et I_c correspondantes.	158
III.28	Signatures ellipsométriques simulés pour les 3 réseaux (A_1 , A_2 et A_3) de la figure III.27 selon les configurations (C_2 , C_3 et C_4) définies dans le tableau III.10.	159
III.29	Cartographie des paramètres géométriques b_1 , b_2 , h et h_r du réseau $E_{h_r}^5$.	161
III.30	Schéma illustrant (a) une déformation du moule dans un processus de nanoimpression et (b) la forme de la résine dans les cavités du moule au fur et à mesure du pressage.	161
III.31	Variation des performances du PMC en fonction du nombre n d'exemples d'entraînement pour différents nombres N_c de neurones cachés.	163
III.32	Erreurs relatives estimées sur les paramètres géométriques b_1 , b_2 et h pour un échantillon simulé en fonction (a) du paramètre b_2 lorsque $b_1 = 30\ \text{nm}$ et $h = 180\ \text{nm}$, et (b) du paramètre h lorsque $b_1 = 30\ \text{nm}$ et $b_2 = 120\ \text{nm}$.	164

III.33	<i>Variations relatives des paramètres géométriques (a) b_1 (b) b_2 et (c) h calculés par le PMC en fonction d'une erreur systématique sur les angles d'incidence θ et d'azimut φ. Les valeurs nominales de b_1, b_2 et h étant respectivement 30nm, 120nm et 180nm.</i>	165
III.34	<i>Erreur relative des paramètres géométriques (a) b_1 (b) b_2 et (c) h calculés par le PMC en fonction de la variation relative de la période pour différentes valeurs du rapport ligne/espace ($R_{l/e}$). Les valeurs nominales de b_1 et h étant respectivement de 25 nm et 180 nm.</i>	167
III.35	<i>Schéma de principe de l'étude concernant l'influence du profil géométrique sur la réponse du PMC.</i>	168
III.36	<i>Profil rectangulaire défini par deux paramètres géométriques, la largeur de la ligne b_L et la hauteur h_L.</i>	168
III.37	<i>(a) Variation relative des paramètres géométriques b_L et h_L par rapport à leurs valeurs nominales, à savoir respectivement 120 nm et 180 nm, en fonction de b_1. (b) Profil rectangulaire obtenu par le PMC ($b_L = 192.2$ nm et $h_L = 159.6$ nm) à partir d'une signature correspondante à un profil trapézoïdal ($b_1 = 50$ nm, $b_2 = 120$ nm et $h = 180$ nm). La période étant fixé à $\Lambda = 565$ nm.</i>	169
III.38	<i>Variation relative des paramètres géométriques b_L et h_L du profil rectangulaire, par rapport à leurs valeurs nominales à savoir $b_L =$ et $h_L =$ en fonction de (a) b lorsque $h = 180$ nm et (b) h lorsque $b = 120$ nm. (c) Exemple d'un profil rectangulaire ($b_L = 80$ nm et $h_L = 155.5$ nm) obtenu par PMC à partir d'une signature ellipsométrique calculée pour un profil sinusoïdal ($b = 120$ nm et $h = 180$ nm).</i>	170
III.39	<i>Variation relative des paramètres géométriques (a) b_1, (b) b_2 et (c) h calculés par le PMC, par rapport à leurs valeurs nominales ($b_1 = 30$ nm, $b_2 = 120$ nm et $h = 180$ nm), en fonction des rayons de courbures bas R_1 et haut R_2. (d) Exemple d'un profil trapézoïdal ($b_1 = 57.7$ nm, $b_2 = 124.9$ nm, $h = 176.9$ nm) obtenu par le PMC à partir d'une signature ellipsométrique correspondante à un profil trapézoïdal arrondi ($b_1 = 30$ nm, $b_2 = 150$ nm, $h = 180$ nm, $R_1 = 70$ nm, $R_2 = 70$ nm.)</i>	171
III.40	<i>(a) Variation relative des paramètres géométriques b, h et h_r calculés par le PMC, par rapport à leurs valeurs initiales ($b = 280$ nm, $h = 250$ nm et $h_r = 40$ nm), en fonction de b_1.</i>	173
III.41	<i>Variation des intensités I_s et I_c en fonction de la longueur d'onde λ pour plusieurs valeurs données de b_2, lorsque (a) $b_1 = 30$ nm et $h = 120$ nm (c) $b_1 = 60$ nm et $h = 220$ nm. Les angles d'incidence et d'azimut sont fixés $\theta = 60^\circ$ et $\varphi = 0^\circ$. (b) et (d) représentent les valeurs absolues des dérivées de I_s et I_c par rapport à b_2 en fonction de la longueur d'onde, calculées pour les valeurs prises respectivement en (a) et en (c).</i>	175
III.42	<i>Intensités I_s et I_c en fonction de la longueur d'onde pour plusieurs valeurs données de h, lorsque (a) $b_1 = 20$ nm et $b_2 = 200$ nm (c) $b_1 = 40$ nm et $b_2 = 90$ nm. Les angles d'incidence et d'azimut sont fixés $\theta = 60^\circ$ et $\varphi = 0^\circ$. (b) et (d) représentent les valeurs absolues des dérivées de I_s et I_c par rapport à h en fonction de λ, calculées pour les valeurs prises respectivement en (a) et en (c).</i>	176

III.43	<i>Evolution de l'erreur quadratique moyenne E_{test} en fonction du nombre d'entrées N.</i>	177
III.44	<i>Evolution de l'erreur quadratique moyenne E_{test} en fonction du nombre de neurones cachés N_c.</i>	177
III.45	<i>(a) Points de mesure choisis pour effectuer une cartographie de l'échantillon E_{hr}^1 sur une dimension de $2\text{ cm} \times 1\text{ cm}$, ainsi que (b) l'image MEB du même échantillon sur une surface de dimension $30\text{ }\mu\text{ m} \times 40\text{ }\mu\text{ m}$.</i>	178
III.46	<i>Cartographie des paramètres géométriques (a) b_1, (b) b_2, (c) h et (d) h_r sur une large surface de l'échantillon ($2\text{ cm} \times 1\text{ cm}$) E_{hr}^1 obtenue par caractérisation neuronale avec signature complète et avec signature optimale sélectionnée par HVS.</i>	179
III.47	<i>Schéma illustrant le principe du classifieur employé pour l'identification du profil géométrique d'un réseau de diffraction à partir de sa signature ellipsométrique mesurée.</i>	180
III.48	<i>Evaluation des performances du classifieur en fonction du nombre d'exemples d'entraînement pour différents nombres de neurones : (a) pourcentage de classification correcte PCC (b) pourcentage de classification incorrecte PCI (c) pourcentage de non classification PNC.</i>	182
III.49	<i>(a) Représentation des deux profils géométriques étudiés : P_1 sinusoïdal (traits continus) tel que $b = 170\text{ nm}$ et $h = 200\text{ nm}$; et P_2 trapézoïdal (traits discontinus) tel que $b_1 = 10\text{ nm}$, $b_2 = 100\text{ nm}$ et $h = 170\text{ nm}$; la période est de 565 nm. (b) Signatures ellipsométriques théoriques correspondantes (I_s, I_c en configuration classique $\varphi = 0^\circ$ tel que l'angle d'incidence $\theta = 60^\circ$) calculées dans le cas de deux réseaux de diffraction en silice.</i>	183
III.50	<i>Sortie s_1 calculée par le classifieur pour un ensemble de 400 échantillons simulés équirépartis pour chaque profil (trapèze avec et sans couche résiduelle h_r). Les 200 premiers échantillons possèdent un profil trapézoïdal simple sans couche et les 200 autres un profil avec une couche résiduelle.</i>	185
III.51	<i>Valeurs des épaisseurs des couches résiduelles h_r correspondantes aux échantillons test de la classe avec couche résiduelle dont la sortie (a) $s_1 < 0,32$ et (b) $s_1 < 0,05$.</i>	186
III.52	<i>Image MEB de l'échantillon montrant la présence d'une couche résiduelle.</i>	187
III.53	<i>Indice de réfraction n et coefficient d'extinction k du silicium.</i>	195
III.54	<i>Indice de réfraction n de la silice.</i>	196
III.55	<i>Indice de réfraction n du verre BK₇.</i>	196
III.56	<i>Indice de réfraction n de la résine utilisée par le LHC, noté RES1.</i>	197
III.57	<i>Indice de réfraction n de la résine utilisée par le LTM, noté RES 2, calculé point par point dans la gamme des longueurs d'onde étudiée.</i>	197

Liste des tableaux

I.1	<i>Nombre d'ordres total (positifs et négatifs) diffractés sur une structure périodique en fonction de la période Λ (nm) et de la longueur d'onde λ_0 (nm) pour une onde incidente avec un angle $\theta_0 = 10^\circ$.</i>	48
I.2	<i>Paramètres A, B, C et h de l'échantillon E_{R_1} déterminés par l'algorithme de Levenberg-Marquardt en considérant un bruit mesuré (σ mesuré) et un bruit fixe (σ cte).</i>	75
II.1	<i>Performances du PMC calculées sur les données test lorsque celui-ci est entraîné sur des données purement théoriques ($\sigma = 0$), et sur des données bruitées ($\sigma = 3.10^{-3}$).</i>	102
II.2	<i>Définition des domaines de variation de l'épaisseur h.</i>	103
II.3	<i>Définition des trois domaines (D_1, D_2 et D_3) de variation des paramètres A, B, C et h.</i>	105
II.4	<i>Erreurs quadratiques moyennes E_{test} calculées sur chacun des paramètres (A, B, C et h) pour les 3 réseaux de neurones entraînés sur les domaines D_1, D_2 et D_3.</i>	105
II.5	<i>Résultats de la caractérisation neuronale pour un échantillon simulé de paramètres optiques et géométriques (A, B, C et h) connus.</i>	109
II.6	<i>Résultats obtenus par la caractérisation neuronale (RN).</i>	110
II.7	<i>Résultats obtenus par optimisation classique (méthode de Levenberg-Marquardt) pour plusieurs épaisseurs initiales h_i; les valeurs initiales de A, B et C sont fixes et respectivement égales 1.6, 1 et 1 proche de la solution finale.</i>	110
II.8	<i>Liste des intensités (I_s, I_c) sélectionnées par la méthode HVS. L'importance estimée de chacune d'elle est donnée en indice (\bullet_1 par exemple correspond à l'intensité la plus influente.)</i>	115
II.9	<i>Caractérisation des paramètres A, B, C et h de la couche de résine par le PMC optimal sélectionné par la méthode HVS (11 entrées et 6 neurones).</i>	115
II.10	<i>Probabilités a posteriori estimées par le réseau de neurones classifieur à partir des mesures ellipsométriques effectuées sur les trois échantillons E_1, E_2 et E_3.</i>	125

III.1 Paramètres géométriques b_1 , b_2 et h et période Λ extraits des images obtenues par MEB et AFM, dans le cas d'un modèle de profil trapézoïdal.	145
III.2 Paramètres géométriques b_1 , b_2 , h et Λ obtenus, selon un profil trapézoïdal symétrique, à partir des 4 PMC notés RN_1 , RN_2 , RN_{31} et RN_{32}	146
III.3 Erreurs quadratiques moyennes calculées pour 100 échantillons test pour les PMC, notés RN_a et RN_b , dédiés à la caractérisation respective de réseaux de profil trapézoïdal symétrique simple (RN_a) et trapézoïdal symétrique arrondi (RN_b).	152
III.4 Paramètres géométriques du profil de l'échantillon en résine déposé sur un substrat de BK_7 estimés par la méthode neuronale selon un profil trapézoïdal symétrique et un profil trapézoïdal symétrique arrondi.	152
III.5 Paramètres géométriques et période du profil de l'échantillon E_{hr}^1 estimé par la méthode neuronale selon un profil trapézoïdal avec une couche résiduelle.	154
III.6 Descriptif des propriétés attendues des réseaux fabriqués par nanoimpression sur une couche de résine photosensible d'épaisseur $h_i = 215$ nm déposée sur un substrat de silicium.	155
III.7 Domaines de variation des paramètres géométriques b_1 , b_2 , h et h_r , notés D_{hr}^2 , D_{hr}^3 , D_{hr}^4 utilisés pour construire le corpus d'apprentissage de chaque PMC.	156
III.8 Paramètres géométriques b_1 , b_2 , h et h_r des échantillons E_{hr}^1 , E_{hr}^2 et E_{hr}^3 résultants de la caractérisation neuronale. Comparaison avec la valeur théorique de h_r obtenue par la loi de conservation du volume de la matière déplacée.	156
III.9 Erreurs quadratiques moyennes E_{test} calculées pour 100 échantillons test pour le PMC dédié à la caractérisation de réseau en silice de période $\Lambda = 565$ nm et de profil trapézoïdal asymétrique.	158
III.10 Différentes signatures ellipsométriques choisies en fonction de la longueur d'onde λ , l'angle d'incidence θ , l'angle d'azimut φ et l'ordre de diffraction.	158
III.11 Erreurs quadratiques moyennes E_{test} calculées sur un ensemble de 200 échantillons test pour les 3 PMC utilisant les signatures C_2 , C_3 et C_4 définies dans le tableau III.10.	160
III.12 Liste des intensités (I_s , I_c) sélectionnées par la méthode HVS. L'importance de chacune d'elle est donnée en indice.	178
III.13 Comparaison entre les paramètres géométriques du profil de réseau, noté E_{hr}^1 , estimés par le PMC avant et après sélection HVS.	178
III.14 Valeurs des sorties s_1 et s_2 calculées par le classifieur entraîné avec $N_c = 5$ pour différents nombres d'exemples d'entraînement n , à partir des signatures ellipsométriques bruitées pour les deux réseaux P_1 et P_2	183
III.15 Valeurs des sorties s_1 et s_2 calculées par le classifieur entraîné avec $N_c = 5$ pour différents nombres d'exemples d'entraînement n , à partir de la signature ellipsométrique mesurée sur le réseau en SiO_2 de période $\Lambda = 565$ nm.	184
III.16 Valeurs des sorties s_1 et s_2 calculées par le classifieur entraîné avec $N_c = 5$ pour différents nombres d'exemples d'entraînement n , pour le réseau E_{hr}^5	187
III.17 Paramètres géométriques du réseau E_{hr}^5 estimés par le PMC dans le cas d'un profil trapézoïdal avec couche résiduelle.	187

Introduction générale

L'industrie de la microélectronique est sans doute l'acteur principal de l'évolution technologique depuis les années 1980. Le thème qui domine ses activités et dans lequel se rassemble le plus d'investissements s'articule autour de la miniaturisation. Cette opération vise à réduire par exemple, la taille d'un transistor qui est à la base de tout circuit électronique intégré. Les méthodes de fabrication utilisées à l'heure actuelle, essentiellement la lithographie, sont suffisamment matures pour assurer une réduction importante de la taille des composants. Malgré cela, les recherches se poursuivent pour aller encore plus loin. D'autres domaines comme l'optique, cherchent à transférer ce savoir faire technologique dans le but de réaliser des composants ouvrant la voie à de nouvelles applications (photoniques, télécommunications optiques, etc...). Toutes les applications envisagées imposent une micro- voir une nano-structuration de ces composants.

Or, il est essentiel que ces avancées soient accompagnées de techniques de caractérisation performantes pour valider les futurs nœuds technologiques, si possible à moindre coût. De plus, le besoin d'un contrôle rapide permettant de suivre en temps réel un processus de fabrication est également recherché. Cette étape devient de plus en plus complexe face à l'augmentation continue de la densité d'intégration des composants.

Les outils de caractérisation usuels, employés généralement pour répondre à ces besoins, sont les méthodes microscopiques. Parmi les méthodes les plus employées, on note la microscopie à force atomique et la microscopie électronique à balayage. La première est limitée par la taille de la pointe qu'elle utilise pour imager la surface de l'échantillon. Malgré les récentes améliorations apportées à cette technique, elle ne constitue pas pour le moment un bon candidat pour la caractérisation des futurs nœuds technologiques. La deuxième est omniprésente au sein de l'industrie, et, est considérée comme une technique de référence pour la caractérisation des dimensions critiques fabriquées. La possibilité de fournir une image réelle de l'échantillon constitue son principal avantage. Par contre, elle s'avère destructive dans certaines conditions d'utilisation et commence à atteindre ses limites en résolution. De plus, ces deux techniques ne permettent qu'une caractérisation très locale, inefficace pour le contrôle des grandes surfaces. En raison de ces limites, le besoin de développer de nouvelles méthodes capables de suivre les progrès technologiques

s'est accru.

Au cours de ces dernières années, l'attention se porte essentiellement sur les méthodes optiques. Ces dernières appartiennent à la famille des méthodes qualifiées d'indirectes. Elles sont basées sur le fait que la signature optique mesurée sur un échantillon donné est fortement dépendante de ses caractéristiques (optiques, géométriques, etc...). Ces méthodes nécessitent deux étapes distinctes : la mesure et le traitement. La première étape consiste à mesurer une signature optique de l'échantillon par différents instruments (ellipsomètre, réflectomètre, etc...). La deuxième étape consiste à déterminer les paramètres caractérisant l'échantillon à partir de la signature mesurée par résolution d'un problème inverse. La plupart des méthodes classiques utilisées dans la littérature pour résoudre ce problème montrent certaines limitations (durée de traitement, efficacité, etc...) et nécessitent davantage de développements. Les métaphores biologiques, comme les réseaux de neurones, suscitent l'attention des chercheurs pour leurs efficacités étonnantes. Ils commencent à s'imposer dans de nombreux domaines de la physique et se présentent comme des outils fortement adaptables au contexte industriel, notamment en métrologie. Le traitement instantané de l'information est le principal avantage des réseaux de neurones par rapport aux méthodes classiques. La durée de caractérisation globale peut ainsi être réduite au seul temps nécessaire à la mesure de la signature optique.

Cette thèse s'intègre parfaitement dans cette thématique de recherche. Son but est d'étudier l'apport des réseaux de neurones dans la caractérisation ellipsométrique, afin de proposer une méthode de caractérisation qui peut être considérée comme une véritable alternative aux méthodes classiques. Nous nous sommes concentrés sur l'étude de deux types de structures, à savoir les couches minces et les réseaux de diffraction. Ce choix est guidé par les nombreuses recherches actuelles visant à réduire les dimensions critiques de ces structures. Le réseau de neurones est employé pour résoudre le problème inverse afin de déterminer les propriétés optiques et géométriques de ces structures à partir d'une mesure ellipsométrique. Dans le cas des réseaux de diffraction, on parle plus particulièrement de diffractométrie ou de scatterométrie puisque la signature mesurée n'est autre que la lumière diffractée par ces structures.

Ce travail de thèse est effectué au sein du laboratoire Dispositifs et Instrumentation Optoélectronique et Microondes (DIOM) de l'Université Jean Monnet de Saint-Etienne dans l'équipe magnéto-optique. Depuis sa création, le laboratoire a acquis un savoir faire dans le domaine de l'optique polarimétrique et l'ellipsométrie [1]. Par ailleurs, de nouvelles compétences en optimisation, et notamment en réseaux de neurones, ont émergé au laboratoire au début de l'année 2004. Ainsi, l'emploi de nouveaux algorithmes de résolution de problème inverse permettant de pallier le manque d'efficacité des méthodes classiques d'optimisation a tout naturellement été envisagé et constitue la base de ce travail de thèse. En vue de poursuivre les activités dans ce domaine, cette équipe vise à développer des méthodes dédiées à la caractérisation de structures optiques complexes comme les structures multicouches et les structures périodiques diffractantes. Ce type de structures fait partie intégrante de plusieurs projets de recherche dirigés par cette équipe, et nécessitent une caractérisation optique et géométrique fiable et efficace.

Cette thèse est le fruit d'une contribution de différents collaborateurs :

- L'équipe Electromagnétisme du Laboratoire LASMEA (Laboratoire des Sciences et Matériaux pour l'Electronique et d'Automatique) UMR 6602 - CNRS de Clermont Ferrand, pour le calcul électromagnétique des structures périodiques diffractantes,

- Le laboratoire Hubert Curien UMR 5516 - CNRS de Saint Etienne pour la fabrication des couches minces de résine et des réseaux de diffraction par lithographie optique ainsi que pour la caractérisation microscopique,
- Le laboratoire LOCEAN (Laboratoire d’Océanographie et du Climat : Expérimentation et Approches Numériques) UMR 7159 - CNRS de Paris pour ses compétences reconnues dans le domaine des réseaux de neurones,
- L’équipe lithographie du laboratoire LTM (Laboratoire des Technologies de la Microélectronique) UMR 5129 - CNRS de Grenoble pour la fabrication des réseaux de diffraction par nanoimpression ainsi que pour leur caractérisation microscopique.

Ce manuscrit s’articule principalement autour de trois chapitres :

Le chapitre I débute par un bref rappel sur la théorie de l’ellipsométrie, en exposant également les principales techniques utilisées pour réaliser une mesure ellipsométrique. Puis, nous aborderons en détail deux types de structures étudiées dans cette thèse et caractérisables par cette technique, à savoir les structures multicouches et les réseaux de diffraction. Pour ces structures, le calcul théorique permettant de simuler les quantités ellipsométriques mesurables, à partir de leurs propriétés optiques et géométriques sera exposé. L’emploi de l’ellipsométrie pour la caractérisation de ces structures nécessite la résolution d’un problème inverse. Son but est d’inférer les paramètres recherchés d’une structure donnée (optiques et/ou géométriques) à partir des mesures ellipsométriques. Le formalisme de résolution du problème inverse en ellipsométrie fera l’objet de la dernière partie du chapitre. Nous passerons en revue les différentes méthodes mathématiques qui peuvent être employées dans ce cadre précis.

Le Chapitre II traite des réseaux de neurones qui s’apparentent à une méthode de régression particulière. Nous nous concentrerons principalement sur une famille de réseaux aux propriétés remarquables appelés perceptrons multicouches qui seront utilisés dans ce travail. Ils seront employés pour résoudre le problème inverse en ellipsométrie, dans le cas d’une caractérisation optique et géométrique de couches minces. Nous illustrerons ainsi la présentation des réseaux de neurones par une application concrète adaptée à ce type de caractérisation. Les intervalles de confiance associés aux différents paramètres seront estimés par un second réseau de neurones. Une méthode de sélection neuronale permettant de choisir les points de mesure les plus pertinents et porteurs d’informations utiles sera ensuite présentée. Elle permettra également de simplifier l’architecture du réseau tout en gardant de bonnes performances. Dans la dernière partie, nous proposerons une autre application qui peut être exploitable en métrologie, à savoir la classification. Elle sera présentée dans le cas particulier de l’identification de la zone d’épaisseur d’une couche mince indépendamment de la phase de caractérisation proprement dite.

Le chapitre III sera totalement consacré à la caractérisation des réseaux de diffraction qui représente toute une activité de recherche regroupée sous le terme de scatterométrie. Le réseau de neurones sera d’abord adapté à résoudre le problème inverse, permettant la détermination du profil géométrique du réseau de diffraction à partir de sa signature ellipsométrique mesurée. Nous étudierons plusieurs types de profils géométriques, et proposerons une méthode pour déterminer la période du réseau de diffraction lorsque celle-ci est mal connue pour des raisons techniques diverses. Ensuite, nous mènerons une étude paramétrique pour comprendre le fonctionnement du réseau de neurones en fonction de différents paramètres qui lui sont intrinsèques et d’autres qualifiés d’externes. Nous examinerons plus particulièrement le comportement du réseau face à des hypothèses plus ou moins vérifiées. La méthode de sélection explicitée précédemment sera également employée

pour réduire la taille de la signature ellipsométrique et simplifier également le réseau de neurones afin d'assurer une caractérisation plus rapide et efficace. Ceci sera mis à profit pour contrôler l'homogénéité des échantillons. Nous montrerons finalement l'apport de la classification en scatterométrie. Dans notre travail, elle sera utilisée pour identifier la forme du profil géométrique d'un réseau de diffraction parmi plusieurs. Elle servira à détecter un défaut dans le processus de fabrication lithographique, à savoir la présence d'une couche résiduelle au fond du masque en résine qui peut être préjudiciable à la phase de gravure ultérieure.

CHAPITRE I

Caractérisation ellipsométrique

L'ellipsométrie est principalement destinée à la caractérisation des propriétés optiques et géométriques des matériaux. C'est une technique de nature optique, non destructive, basée sur l'analyse de l'état de polarisation d'une onde lumineuse, après réflexion sur une surface plane, décrit par les deux angles ellipsométriques ψ et Δ . Elle s'est trouvée largement exploitée dans les laboratoires de recherche depuis son apparition il y a plus d'un siècle. Son utilisation au sein du milieu industriel reste relativement récente, mais devient de plus en plus remarquable au cours de ces dernières années. Elle est devenue l'une des méthodes références pour la détermination des épaisseurs de couches minces pour une large gamme de matériaux visant diverses applications. L'ellipsométrie est une méthode qualifiée d'indirecte, et nécessite deux étapes distinctes : la mesure et le traitement. En effet, les mesures doivent être traitées mathématiquement afin de déterminer les propriétés optiques ou géométriques recherchées.

Le but de la première partie de ce chapitre est de fournir au lecteur les bases théoriques de l'ellipsométrie. Après avoir revu les principales méthodes permettant de réaliser une mesure ellipsométrique, nous présenterons en détail le dispositif expérimental utilisé dans le cadre de cette thèse, à savoir un ellipsomètre spectroscopique à modulation de phase.

Ensuite, nous exposerons les différents types de structures qui seront caractérisées, à savoir les structures multicouches et les réseaux de diffraction. Nous détaillerons ainsi le fondement mathématique, permettant le calcul théorique direct des angles ellipsométriques ψ et Δ à partir des propriétés optiques et géométriques de ces structures, comme par exemple l'indice de réfraction et l'épaisseur. Dans le cas des structures multicouches, ce calcul peut être achevé analytiquement en employant un formalisme matriciel approprié ainsi que les lois classiques de l'optique géométrique. Dans le cas plus complexe des réseaux de diffraction, ce calcul ne possède pas de solution analytique. Les différentes méthodes développées dans la littérature seront revues, en mettant plus particulièrement l'accent sur celle que nous emploierons par la suite, à savoir la méthode modale par dé-

veloppement en série de Fourier dite MMMFE (de l'anglais Multilayer Modal Method by Fourier Expansion).

Dans la dernière partie de ce chapitre, nous présenterons brièvement les différentes méthodes mathématiques qui peuvent être employées pour déterminer les paramètres optiques et géométriques à partir de données ellipsométriques.

I.1 Ellipsométrie classique

L'objectif principal de cette partie est de présenter succinctement l'ellipsométrie. Après un bref historique et une présentation rapide de différentes applications caractéristiques, nous exposerons les fondements théoriques. Nous finirons par rappeler les principales méthodes utilisées à l'heure actuelle pour la réalisation d'une mesure ellipsométrique.

I.1.1 Histoire et applications

La première étude ellipsométrique reportée dans la littérature date de 1887 quand le physicien allemand Paul Drude établit les équations fondamentales [2] qui constituent la base de l'ellipsométrie actuelle. Le principe consiste à mesurer le changement de l'état de polarisation d'une onde lumineuse après réflexion sur une surface plane. Au delà de sa contribution théorique, il a aussi réalisé les premières études expérimentales [3, 4] sur des matériaux absorbants et transparents.

Pendant les 75 ans qui ont suivi le travail pionnier de Paul Drude, les études ellipsométriques réalisées furent peu nombreuses. Cependant l'une des plus notables fut celle de Rothen [5] qui a donné à cette technique le nom actuel d'"ellipsométrie" pour la distinguer de la mesure polarimétrique.

Ce n'est que dans les années 1960 que l'ellipsométrie a connu une renaissance grâce à la présence des ordinateurs pour le calcul numérique des données de plus en plus complexes. Depuis, elle est devenue l'une des techniques les plus utilisées pour la détermination des propriétés optiques (indice de réfraction et coefficient d'extinction) et géométriques (épaisseur, rugosité de surface, etc...) d'une large gamme de matériaux.

Aujourd'hui, cette technique est employée dans de nombreux domaines, aussi bien dans les laboratoires de recherche que dans le milieu industriel.

En microélectronique, la lithographie utilise des longueurs d'onde d'insolation des résines photosensibles de plus en plus courtes pour augmenter la densité d'intégration des composants. L'ellipsométrie s'avère alors utile notamment pour déterminer l'indice optique de la résine à ces longueurs d'onde [6], ainsi que d'autres paramètres comme la température de transition vitreuse [7] nécessaire par exemple dans un procédé de nano-impression thermique. Cette technique a également servi pour la détermination des propriétés diélectriques de nouveaux matériaux, comme les nanoparticules de germanium [8] présentant un intérêt grandissant pour l'intégration des composants optoélectroniques.

Dans la technologie des écrans plats de type TFT-LCD (Thin Film Transistor-Liquid Crystal Display), l'ellipsométrie est utilisée pour déterminer les performances (luminosité et contraste) du système [9], directement liées aux propriétés optiques des matériaux. Elle permet aussi d'effectuer des cartographies afin de contrôler l'homogénéité de la surface des écrans devenant de plus en plus large. De même, la caractérisation des structures complexes, comme les multicouches pour la réalisation de miroirs à haute réflectivité est

rendue possible grâce à l'ellipsométrie (23 couches successives de SiO_2-TiO_2 déposées sur un substrat [10]).

En magnéto-optique, on peut caractériser avec précision la biréfringence et le dichroïsme de certains liquides magnétiques [11], ainsi que les propriétés optiques de matériaux sol-gel [12].

En biologie, l'ellipsométrie permet de détecter des couches ultraminces de biomolécules sur des surfaces solides [13] ainsi que la concentration de ces molécules [14], ouvrant de nouvelles voies vers des applications de dosage en biochimie et biomédecine.

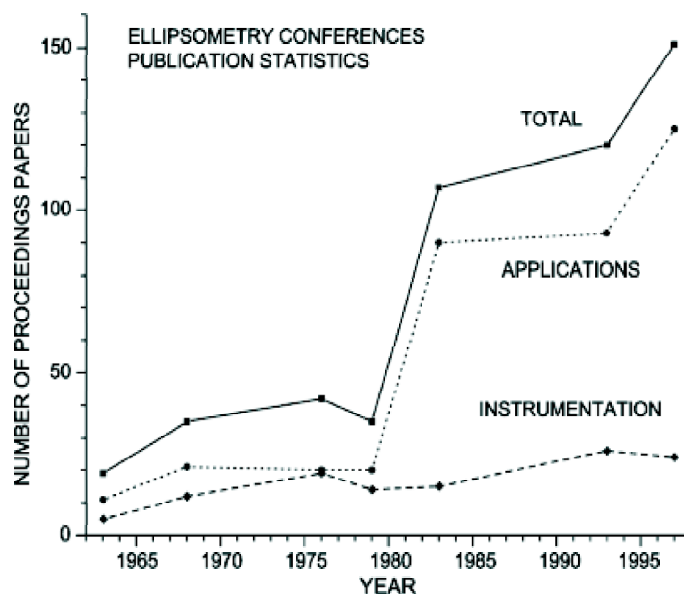


FIG. I.1: Nombre de publications issues des précédentes conférences sur l'ellipsométrie, concernant deux catégories : les applications et l'instrumentation (ref : D. E. Aspnes, *Thin Solid Films* 455-456 (2004)).

Enfin, l'ellipsométrie a démontré sa capacité de contrôler en temps réel un processus de fabrication, comme par exemple le suivi de la croissance de couches minces [15, 16], permettant ainsi de comprendre et d'ajuster si nécessaire les étapes de fabrication industrielle.

La liste des applications de l'ellipsométrie présentée ici est loin d'être exhaustive. La figure I.1 montre le nombre des publications issues des différentes conférences sur l'ellipsométrie [17] depuis les années 1960 jusqu'au début des années 2000. Bien que l'aspect instrumental de la technique soit presque stable, on voit clairement que le nombre d'applications est en perpétuelle augmentation ces dernières années et ne cesse pas d'évoluer.

Après cette brève introduction à l'ellipsométrie, nous allons à présent nous intéresser au principe de fonctionnement de la méthode.

I.1.2 Fondements théoriques

Considérons le cas simple d'une onde plane arrivant sur une surface plane, sous un angle d'incidence quelconque θ_0 , séparant deux milieux, supposés homogènes et isotropes, notés 0 et 1 d'indice respectif N_0 et N_1 (figure I.2). Une partie de l'onde incidente est

ainsi réfléchi dans le même milieu, une autre partie est transmise (ou absorbée) à travers la surface.

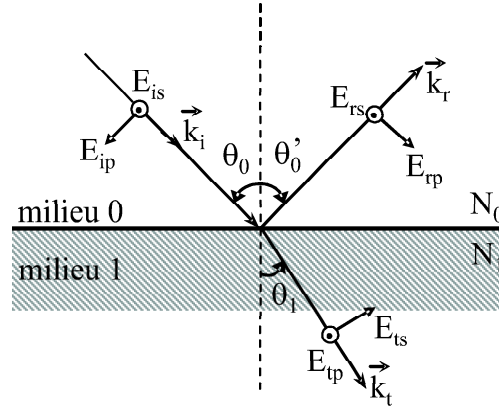


FIG. I.2: Réflexion et transmission d'une onde plane incidente, avec un angle θ_0 , sur une surface plane séparant deux milieux 0 et 1 d'indice respectif N_0 et N_1 . On montre sur la figure les composantes s et p du champ électrique des ondes incidente (E_i), réfléchie (E_r) et transmise (E_t) ainsi que les vecteurs d'onde associés (\vec{k}_i , \vec{k}_r et \vec{k}_t).

Les phénomènes de réflexion et de transmission sont régis par les lois de Descartes :

$$\theta_0 = -\theta'_0 \quad (\text{I.1})$$

$$N_0 \sin \theta_0 = N_1 \sin \theta_1 \quad (\text{I.2})$$

Le milieu 0 est en général le milieu ambiant ($N_0 = 1$) et l'indice du milieu 1 peut être écrit sous sa forme complexe $N_1 = n_1 + jk_1$, telle que n_1 est l'indice réel de réfraction et k_1 le coefficient d'extinction du milieu.

Les directions de propagation des ondes incidente, réfléchie et transmise données respectivement par les vecteurs d'onde \vec{k}_i , \vec{k}_r et \vec{k}_t sont contenues dans le plan d'incidence. Ce plan est défini comme étant le plan formé par la normale à la surface et la direction incidente \vec{k}_i . On désigne par \vec{E}_p la composante vectorielle parallèle du champ électrique (p pour onde transverse magnétique ou TM) et \vec{E}_s la composante vectorielle perpendiculaire (s pour onde transverse électrique ou TE), par rapport au plan d'incidence. Les champs électriques des ondes incidente (\vec{E}_i), réfléchie (\vec{E}_r) et transmise (\vec{E}_t) peuvent être définis par les équations suivantes :

$$\vec{E}_i = \vec{E}_{ip} + \vec{E}_{is} \quad (\text{I.3})$$

$$\vec{E}_r = r_p \vec{E}_{ip} + r_s \vec{E}_{is} \quad (\text{I.4})$$

$$\vec{E}_t = t_p \vec{E}_{ip} + t_s \vec{E}_{is} \quad (\text{I.5})$$

avec :

- r_p , r_s : coefficients de réflexion en amplitude dans le cas d'une polarisation respectivement p et s ,
- t_p , t_s : coefficients de transmission en amplitude dans le cas d'une polarisation respectivement p et s .

Ces coefficients, nommés coefficients de Fresnel, sont donnés par les équations suivantes (la notation 01 correspond au cas où l'onde est incidente dans le milieu 0 et transmise

dans le milieu 1) :

$$r_{01p} = r_p = \frac{E_{rp}}{E_{ip}} = \frac{N_1 \cos \theta_0 - N_0 \cos \theta_1}{N_1 \cos \theta_0 + N_0 \cos \theta_1} \quad (\text{I.6})$$

$$r_{01s} = r_s = \frac{E_{rs}}{E_{is}} = \frac{N_0 \cos \theta_0 - N_1 \cos \theta_1}{N_0 \cos \theta_0 + N_1 \cos \theta_1} \quad (\text{I.7})$$

$$t_{01p} = t_p = \frac{E_{tp}}{E_{ip}} = \frac{2N_1 \cos \theta_0}{N_1 \cos \theta_0 + N_0 \cos \theta_1} \quad (\text{I.8})$$

$$t_{01s} = t_s = \frac{E_{ts}}{E_{is}} = \frac{2N_0 \cos \theta_0}{N_0 \cos \theta_0 + N_1 \cos \theta_1} \quad (\text{I.9})$$

Remarque : Dans le cas où l'onde est incidente du milieu 1 vers le milieu 0, les coefficients de réflexion et de transmission notés r_{10} et t_{10} peuvent être calculés en remplaçant simplement N_0 par N_1 et vice versa dans les équations I.6-I.9, pour obtenir ainsi :

$$r_{10} = -r_{01} \quad (\text{I.10})$$

$$t_{10} = (1 - r_{01}^2) / t_{01} \quad (\text{I.11})$$

Pour une onde incidente donnée, l'amplitude et la polarisation des ondes, réfléchie et transmise, sont déterminées par la continuité des composantes tangentielles des champs électriques et magnétiques à travers la surface. Ainsi, il est facile de constater que si l'onde incidente est linéairement polarisée TE ou TM , les ondes, réfléchie et transmise, auront le même état de polarisation.

En se limitant au cas de la réflexion (même raisonnement en transmission), les coefficients r_p et r_s , qui sont de nature complexe, peuvent s'écrire :

$$r_p = \frac{E_{rp}}{E_{ip}} = |r_p| e^{j\delta_p} \quad (\text{I.12})$$

$$r_s = \frac{E_{rs}}{E_{is}} = |r_s| e^{j\delta_s} \quad (\text{I.13})$$

avec :

- $|r_p|$ et $|r_s|$: modules des coefficients de réflexion représentant la modification apportée en amplitude à la composante p et s du champ incident,
- δ_p et δ_s : retards de phase introduits par la réflexion dans le cas d'une polarisation p et s .

En pratique, la quantité mesurée est le rapport de ces deux coefficients. Elle est définie comme étant l'équation fondamentale de l'ellipsométrie et s'exprime sous la forme :

$$\rho = \frac{r_p}{r_s} = \tan \psi \exp(j\Delta) \quad (\text{I.14})$$

soit par identification :

$$\psi = \arctan \left| \frac{r_p}{r_s} \right|$$

$$\Delta = \delta_p - \delta_s$$
(I.15)

Ces angles peuvent varier dans les domaines suivants :

$$0 < \psi < 90^\circ \quad 0 < \Delta < 180^\circ$$

Il est évident que le calcul théorique des angles ellipsométriques I.14 nécessite la connaissance des coefficients de réflexion r_p et r_s dépendants de la structure étudiée.

Bien entendu, ces coefficients ne dépendent pas seulement de l'angle d'incidence θ_0 , mais aussi de la longueur d'onde λ de la lumière incidente par la dispersion intrinsèque du milieu 1 ($N_1(\lambda)$). Une fois que les coefficients de réflexion r_p et r_s sont calculés, les angles ellipsométriques ψ et Δ peuvent être directement déterminés à partir de l'équation fondamentale I.14.

Considérons par exemple deux cas analogues à celui de la figure I.2 (pour un milieu 1 transparent (silice) et absorbant (silicium)) pour une longueur d'onde $\lambda = 350 \text{ nm}$. Les lois de dispersions de ces matériaux sont bien entendu supposées connues et données dans l'annexe A. La figure I.3 montre les coefficients ψ et Δ en fonction de l'angle d'incidence pour les deux types de matériaux considérés.

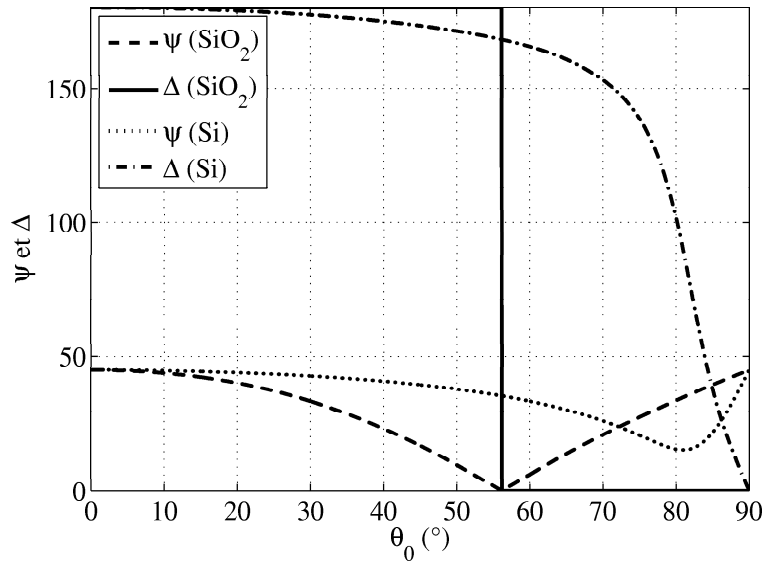


FIG. I.3: Angles ellipsométriques ψ et Δ en fonction de l'angle d'incidence θ_0 dans le cas des interfaces air-silice et air-silicium pour une longueur d'onde $\lambda = 350 \text{ nm}$.

A un angle d'incidence particulier, ψ atteint une valeur minimale et Δ passe d'une valeur de 180° à 0° . Dans le cas d'un matériau transparent, ψ s'annule et la variation de Δ est discontinue à cet angle, appelé communément *Angle de Brewster* θ_b ($\theta_b = 56,15^\circ$). A cet angle particulier, une onde incidente de polarisation quelconque se trouve polarisée *s* ($\psi = 0$ donc $r_p = 0$) après réflexion sur la surface. Pour cette raison, on parle également d'*angle de polarisation*. Il est déterminé à partir de l'équation I.6 et I.1, en posant $r_p = 0$:

$$\theta_b = \arctan \frac{N_1}{N_0} \quad (\text{I.16})$$

Pour les matériaux absorbants, cet angle est appelé *angle de pseudo-Brewster* pour lequel la valeur de ψ ne s'annule pas mais atteint un minimum qui dépend de la valeur du coefficient d'extinction k_1 du matériau à la longueur d'onde d'étude. Les variations de Δ sont cette fois-ci continues. L'angle θ_b est toujours calculé par la même formule I.16, et vaut $\theta_b = 81^\circ$ dans le cas du silicium, vérifié également sur la figure I.3.

On doit finalement noter que la variation de ψ et Δ en fonction de l'angle d'incidence montrent les zones favorables de mesure. Ainsi, au voisinage de l'angle de Brewster la sensibilité de mesure de ψ et Δ est maximale.

Notons également que lorsqu'une onde polarisée linéairement se réfléchit sur une surface, sa polarisation devient elliptique. Les caractéristiques de cette ellipse dépendent essentiellement des caractéristiques de cette surface (par exemple n_1 , k_1) qui sont contenues dans les angles ellipsométriques ψ et Δ . L'objectif de l'ellipsométrie est de remonter aux caractéristiques de cette surface par une analyse de cet état de polarisation émergent.

Pour la mesure des angles ellipsométriques déjà définis, nous allons voir maintenant les différentes techniques largement employées dans la littérature.

I.1.3 Différentes techniques de mesure

La configuration ellipsométrique de base permettant de mesurer ψ et Δ se compose d'une source, d'un polariseur, (éventuellement d'un compensateur), de l'échantillon à étudier, d'un analyseur et d'un détecteur (figure I.4). On parle de montage PCSA (de l'anglais Polarizer-Compensator-Surface-Analyzer). Il existe plusieurs façons de réaliser un système de mesure ellipsométrique à partir de cette configuration.

Remarque : Notons que le compensateur (en fait une lame quart d'onde) peut également être placé sur le bras détecteur entre l'échantillon et l'analyseur. La configuration sera alors notée PSCA (de l'anglais Polarizer-Surface-Compensator-Analyzer).

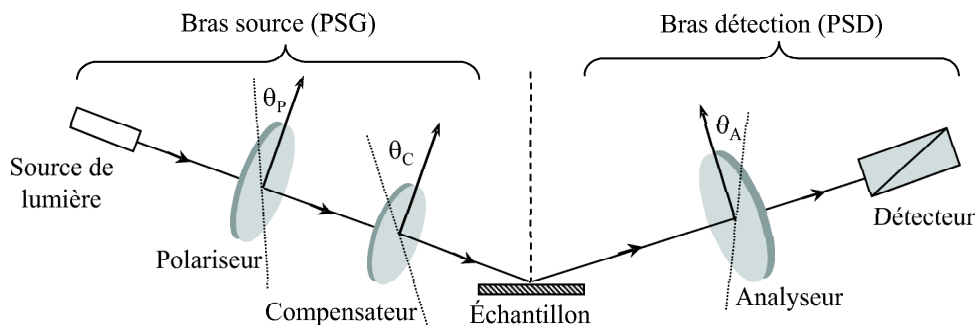


FIG. I.4: Montage ellipsométrique de base (PCSA) comportant un polariseur, un compensateur, l'échantillon et l'analyseur. Les pointillés représentent les axes de transmission de chaque composant optique orienté par rapport au plan d'incidence (représenté par le plan formé par le faisceau lumineux et la normale à l'échantillon) suivant les angles θ_P , θ_C et θ_A .

Chaque élément optique (polariseur, compensateur et analyseur) est désigné par l'angle que fait son axe de transmission avec le plan d'incidence (respectivement θ_P , θ_C et θ_A).

Dans un système ellipsométrique classique (figure I.4), les différents éléments sont arrangés selon un bras source (ou PSG de l'anglais Polarization Stage Generator) et un bras de détection (ou PSD de l'anglais Polarization State Detector).

Les méthodes ellipsométriques peuvent être classées en deux grandes catégories : polarimétriques et photométriques. Dans le premier cas, c'est une mesure basée sur la recherche des angles (θ_P , θ_C et θ_A) des différents éléments annulant l'intensité au niveau du détecteur. C'est l'*ellipsométrie à annulation* dite *méthode de zéro*. Dans le cas des systèmes photométriques, on réalise une mesure d'intensité modulée ; soit à l'aide d'un élément tournant soit en introduisant un modulateur ; tout en gardant fixe les autres éléments du montage.

I.1.3.1 Ellipsomètre à annulation

L'une des méthodes ellipsométriques les plus anciennes qui fut réalisée en mono longueur d'onde, est l'ellipsométrie à extinction [18]. L'arrangement des composants optiques est tel qu'il est montré dans la figure I.4.

Le concept de cet ellipsomètre est simple. La lumière issue de la source est polarisée linéairement après traversée du polariseur. Ensuite, l'état de polarisation devient, a priori, elliptique au delà du compensateur. Suite à sa réflexion sur l'échantillon, supposé plan et isotrope, l'onde reste en général de polarisation elliptique.

Il est possible de trouver une position particulière de θ_A et θ_C telle que, après réflexion sur l'échantillon, la polarisation de l'onde redevienne linéaire. L'analyseur, à son tour, est orienté de façon à être croisé avec la polarisation linéaire déjà produite, ce qui conduit à l'extinction du faisceau au niveau du détecteur. Le zéro de l'intensité est ainsi obtenu pour un couple (θ_P , θ_A) approprié. L'échantillon joue ici un rôle symétrique à la lame quart d'onde ; en effet, une telle lame transforme une polarisation rectiligne en polarisation elliptique, alors que l'échantillon réalise l'opération inverse après réflexion. En pratique, l'angle θ_C du compensateur est fixé à 45° .

Connaissant les angles θ_P et θ_A annulant l'intensité, les angles ellipsométriques ψ et Δ sont déterminés à partir de l'équation suivante [18, 19] :

$$\tan \psi \exp(j\Delta) = -\tan \theta_A \frac{\tan \theta_C - \tan(\theta_P - \theta_C)}{1 + j \tan \theta_C \tan(\theta_P - \theta_C)} \quad (\text{I.17})$$

Le calcul des angles ψ et Δ est ainsi exempt d'ambiguïté. Néanmoins, cette technique reste lente malgré le développement des microprocesseurs et l'automatisation des différents composants, puisqu'elle nécessite toujours la recherche fastidieuse du minimum d'intensité.

Bien qu'elle soit basée sur une recherche angulaire, cette méthode est précise et possède peu d'erreurs systématiques. En revanche, cette précision dépend fortement du bruit de fond du détecteur, lié à la mesure du minimum d'intensité. De plus, l'utilisation spectroscopique de cette méthode est délicate car elle nécessite un compensateur "achromatique" devant être étalonné pour chaque longueur d'onde. Il est cependant possible d'augmenter la précision de mesure en prenant plusieurs valeurs de θ_C , donc plusieurs couples (θ_P , θ_A), pour calculer la valeurs des angles ψ et Δ .

I.1.3.2 Ellipsomètre à modulation par élément tournant

L'avènement de l'automatisation a permis de mettre en place la seconde famille d'ellipsomètre incorporant une modulation. Cette modulation peut être produite par la rotation du polariseur (en anglais RPE pour Rotating Polarizer Element), de l'analyseur (RAE pour Rotating Analyzer Element) ou du compensateur (RCE pour Rotating Compensator Element). Ils suivent le même principe que le schéma de la figure I.4, exception faite pour les deux premières méthodes citées qui ont été envisagées à la base sans compensateur.

La grande différence avec l'ellipsomètre à extinction est qu'un élément est soumis à une rotation uniforme lorsque les autres sont maintenus à angle fixe. L'effet de cette rotation est de moduler en polarisation le faisceau lumineux. Le signal détecté étant mesuré en fonction du temps, une analyse mathématique (généralement une transformée de Fourier) est nécessaire pour déterminer les angles ellipsométriques ψ et Δ .

Par exemple, l'intensité détectée à la sortie d'un système RPE prend la forme suivante (le traitement est identique dans les deux autres cas) :

$$I = I_N(\alpha \cos 2\theta_P + \beta \sin 2\theta_P + 1) \quad (\text{I.18})$$

avec

$$\alpha = \frac{\tan^2 \psi - \tan^2 \theta_A}{\tan^2 \psi + \tan^2 \theta_A} \quad (\text{I.19})$$

$$\beta = 2 \cos \Delta \frac{\tan \psi \tan \theta_A}{\tan^2 \psi + \tan^2 \theta_A} \quad (\text{I.20})$$

$$I_N = \frac{|r_s|^2 |\vec{E}_0|^2}{2} \cos^2 \theta_A (\tan^2 \psi + \tan^2 \theta_A) \quad (\text{I.21})$$

où $|\vec{E}_0|^2$ est l'intensité initiale du faisceau incident noté I_0 .

En pratique, les paramètres α et β sont mesurés par détection synchrone (θ_P étant proportionnel à $\cos(\omega t)$ où ω est la vitesse de rotation du polariseur). Notons que seule I_N dépend de l'intensité initiale du faisceau incident. Ainsi, les angles ψ et Δ peuvent être déduits en fonction de θ_A , α et β :

$$\tan \psi = \sqrt{\frac{1+\alpha}{1-\alpha}} \tan \theta_A \quad (\text{I.22})$$

$$\cos \Delta = \frac{\beta}{\sqrt{1-\alpha^2}} \quad (\text{I.23})$$

Ce sont les ellipsomètres à analyseur tournant [20, 21] qui ont reçu le plus d'attention au début de la période d'automatisation. Néanmoins, notons que l'erreur expérimentale, lors de la mesure de Δ , est proportionnelle à $1/\sin \Delta$ [22]. Ceci conduit à une mauvaise précision lorsque la valeur de Δ est proche de 0° ou 180° . Suite à cette limitation, plusieurs études ont montré [23, 24] que ce problème peut être surmonté en introduisant un compensateur (par exemple une lame quart d'onde) dans le montage ellipsométrique précédent. Ceci permet de modifier artificiellement la valeur de Δ pour sortir de cette zone critique.

Le montage à polariseur tournant est identique au précédent, avec cette fois-ci un analyseur fixe. Cette configuration nécessite une source dont la polarisation est parfaitement connue. Les montages à compensateur tournant [25] ont été également étudiés par plusieurs auteurs [26, 27]. Ils ont permis de s'affranchir des limitations des RPE et RAE. Toutefois, ce type d'ellipsomètre nécessite un étalonnage spectral du compensateur produisant ainsi des erreurs systématiques nuisibles pour la mesure.

La communauté scientifique continue toujours à concevoir et améliorer le montage à élément tournant, comme par exemple le PRPCSE [28] (de l'anglais Polarizer Rotating-Polarizer Compensator Spectroscopic Ellipsometry) adapté surtout pour la caractérisation des couches transparentes.

L'ellipsomètre à élément tournant est facile à mettre en œuvre et permet une acquisition automatique des données ellipsométriques d'une façon plus rapide que l'ellipsomètre à extinction. L'un des inconvénients majeurs des montages à élément tournant est généralement le problème d'alignement occasionnant des erreurs systématiques. De même, l'introduction d'un compensateur rend plus complexe l'utilisation spectroscopique du dispositif.

I.1.3.3 Ellipsomètre à modulation de phase

L'ellipsomètre à modulation de phase ou PME (de l'anglais Phase Modulation Ellipsometry) est une troisième approche introduite en 1969 par Jaspersen and Schnatterley [29]. L'idée était d'introduire un modulateur photoélastique permettant de moduler la polarisation de la lumière sans aucun mouvement mécanique.

Ce composant est constitué d'un barreau parallélépipédique de matériau transparent, souvent en silice, solidaire d'un élément piézoélectrique oscillant à une fréquence f élevée d'environ 50 kHz utilisée dans la plupart des dispositifs commercialisés. La contrainte sinusoïdale introduite par le matériau piézoélectrique et appliquée selon un axe donné, provoque une anisotropie optique dans le barreau dépendant du temps. L'effet de cette anisotropie est de générer une différence de phase $\delta(t)$ périodique entre les composantes orthogonales des axes propres du barreau. Une onde incidente polarisée linéairement dans une direction différente de celle de l'axe de modulation, se trouve ainsi polarisée elliptiquement à la sortie.

Cette différence de phase prend la forme suivante :

$$\delta(t) = \delta_0 \sin(\omega t) \quad (\text{I.24})$$

avec :

- ω : pulsation angulaire donnée par $2\pi f$.
- δ_0 : amplitude maximale de modulation qui est égale à $2\pi d(n_{1max} - n_0)/\lambda$, où n_1 ($n_{1min} < n_1 < n_{1max}$) et n_0 sont les indices de réfraction suivant chaque axe propre du barreau et d est l'épaisseur de celui-ci.

Le modulateur est souvent placé après le polariseur (à la place du compensateur dans la figure I.4). En général, le polariseur et le modulateur sont rendus solidaires de telle manière que l'angle $\theta_P - \theta_M$ soit fixé à $\pm 45^\circ$, (θ_M étant l'angle du modulateur). Dans ce cas, l'état de polarisation résultant du modulateur est elliptique, avec une ellipticité dite dynamique, c'est à dire l'état de polarisation oscille entre linéaire et elliptique. Ainsi cet

assemblage polariseur/modulateur permet d'engendrer tous les états de polarisation au cours d'un cycle complet de modulation.

L'absence d'éléments tournants dans le dispositif expérimental supprime les erreurs systématiques dues aux défauts d'alignement. Par contre, un système de régulation est nécessaire pour stabiliser la température du modulateur, ainsi qu'un étalonnage de celui-ci lors d'une utilisation en mode spectroscopique [30, 31]. L'acquisition et le traitement du signal à une fréquence compatible avec celle de la modulation, nécessitent également une électronique performante capable d'assurer cette tâche.

L'appareil de mesure utilisé dans ce travail est un ellipsomètre spectroscopique à modulation de phase (UVISEL de Jobin Yvon), représenté schématiquement sur la figure I.5.

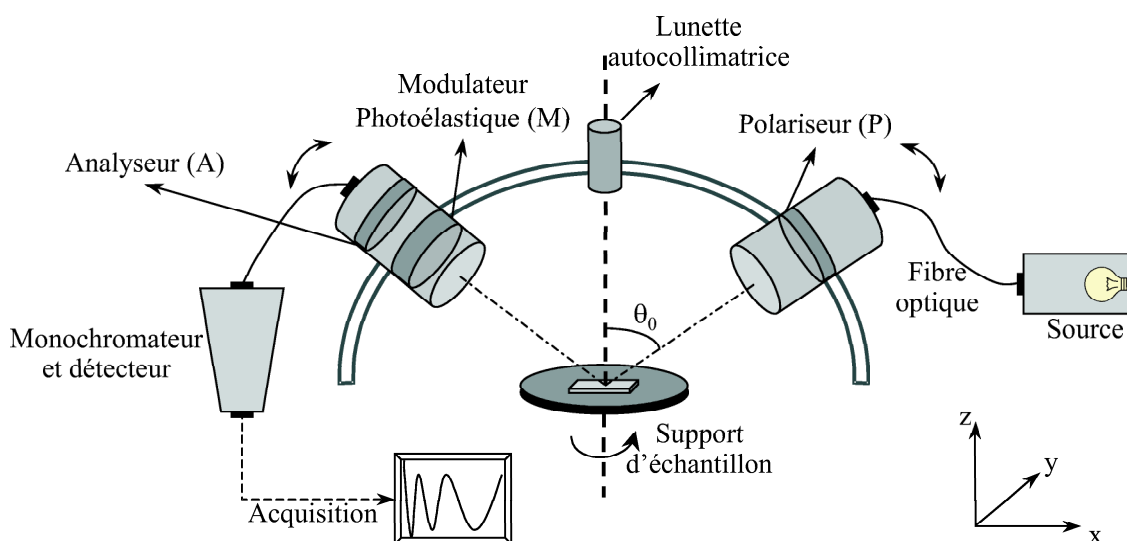


FIG. I.5: Représentation schématique de l'ellipsomètre spectroscopique à modulation de phase (UVISEL de Jobin Yvon).

Il faut noter que sur cet appareil le modulateur se situe sur le bras de détection (PSD) entre l'analyseur et l'échantillon. Le bras source (PSG) contient seulement le polariseur. Les deux bras peuvent être orientés par rotation sur un goniomètre. L'angle d'incidence θ_0 est défini par l'axe optique du bras source et la normale à la surface de l'échantillon. La source utilisée est une lampe Xénon d'une puissance de 75 Watts avec une alimentation stabilisée. Elle assure une large gamme spectrale qui s'étend de l'ultraviolet au proche infrarouge (260 nm - 2100 nm).

Le goniomètre nous permet d'avoir un angle d'incidence pouvant varier automatiquement de 40° à 90° avec une précision de $\pm 0.01^\circ$. L'échantillon est maintenu par aspiration sur un support automatisé permettant d'effectuer soit des déplacements en x et y soit une rotation de 360° autour de la normale. L'horizontalité de la surface est contrôlée à l'aide d'une lunette autocollimatrice. Le polariseur et l'analyseur utilisés sont en calcite (avec un coefficient d'extinction de l'ordre de 10^{-5}) montés sur des platines de rotation commandées par des moteurs pas à pas de résolution $\pm 0.05^\circ$.

L'axe du modulateur photoélastique est repéré par l'angle θ_M . Il est monté sur une platine de rotation automatique et maintenu en température. Il est rendu solidaire à l'analyseur de sorte que $\theta_A - \theta_M$ soit fixé à 45° .

Le système de détection combine l'utilisation d'un monochromateur double sortie à

haute résolution et deux détecteurs. Le monochromateur est équipé d'une tourelle motorisée à deux réseaux de diffraction afin de balayer le domaine de l'UV visible au proche infra rouge. Le premier réseau, de période $0,83 \mu\text{m}$, couvre un domaine spectral qui s'étend de 260 à 830 nm, avec une dispersion linéaire de 2 nm/mm. Le second, de période $1,66 \mu\text{m}$, assure le domaine allant de 800 à 2100 nm, avec une dispersion linéaire de 4 nm/mm. Le premier détecteur est un photomultiplicateur, le second est une photodiode de type InGaAs (Indium-Galium-Arsenide).

La propagation de la lumière à travers un dispositif ellipsométrique peut facilement être décrite par le formalisme de Jones. En effet, l'étude de la propagation de l'état de polarisation d'une onde lumineuse à travers différents éléments optiques se réduit à un simple calcul matriciel.

Pour cela, on définit le repère fixe ($Oxyz$) tel que Oz est la direction de propagation de l'onde lumineuse, Ox et Oy étant contenus dans le plan perpendiculaire à Oz .

La matrice de transfert d'un polariseur (ou analyseur) rectiligne, exprimée dans son repère propre et dont la direction de polarisation est parallèle à l'axe Ox est définie par :

$$P = A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (\text{I.25})$$

Introduisons également la matrice rotation notée $R(\alpha)$, matrice de changement de base entre le repère fixe et un repère (Ox' , Oy') orienté d'un angle α avec Ox tel que $(Ox, Ox') = \alpha$:

$$R(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \quad (\text{I.26})$$

L'action du modulateur et de l'échantillon sont représentés respectivement par les matrices de Jones M (exprimée dans son repère propre) et S (exprimée dans le repère Oxy) suivantes :

$$M = \begin{bmatrix} e^{j\delta(t)} & 0 \\ 0 & 1 \end{bmatrix} \quad ; \quad S = \begin{bmatrix} r_p & 0 \\ 0 & r_s \end{bmatrix} \quad (\text{I.27})$$

Pour une configuration des angles θ_P , θ_M et θ_A quelconque, le vecteur champ électrique \vec{E}_d au niveau du détecteur, exprimé dans le plan fixe, est obtenu par le produit matriciel suivant :

$$\vec{E}_d = [A.R(\theta_A).R(-\theta_M).M.R(\theta_M).S.R(-\theta_P).P] \vec{E}_i \quad (\text{I.28})$$

Après transformation, l'intensité du champ électrique peut se mettre sous la forme :

$$I(t) = E_d E_d^* = \left| \vec{E}_d \right|^2 = B [I_N + I_s \sin \delta(t) + I_c \cos \delta(t)] \quad (\text{I.29})$$

avec

$$I_N = 1 - \cos 2\psi \cos 2\theta_P + \cos[2(\theta_A - \theta_M)] \cos[2\theta_M(\cos 2\theta_P - \cos 2\psi)] \\ + \sin 2\theta_P \cos \Delta \cos[2(\theta_A - \theta_M)] \sin 2\psi \sin 2\theta_M \quad (\text{I.30})$$

$$I_c = \sin[2(\theta_A - \theta_M)] [\sin 2\theta_M(\cos 2\psi - \cos 2\theta_P) + \sin 2\theta_P \cos 2\theta_M \sin 2\psi \cos \Delta] \quad (\text{I.31})$$

$$I_s = \sin[2(\theta_A - \theta_M)] \sin 2\theta_P \sin 2\psi \sin \Delta \quad (\text{I.32})$$

$$B = \frac{I_0}{4(r_p^2 + r_s^2)} \quad (\text{I.33})$$

avec $I_0 = |\vec{E}_0|^2$ est l'intensité initiale du faisceau incident.

La différence de phase $\delta(t)$ étant une fonction sinusoïdale, les deux fonctions $\sin \delta(t)$ et $\cos \delta(t)$ peuvent être exprimées par un développement en série de Fourier sur la base des fonctions de Bessel de première espèce avec comme argument l'amplitude de modulation δ_0 (équation I.24) :

$$\begin{aligned} \sin \delta(t) &= 2 \sum_{m=0}^{\infty} J_{2m+1}(\delta_0) \sin[(2m+1)\omega t] \\ &= 2J_1(\delta_0) \sin(\omega t) \\ \cos \delta(t) &= J_0(\delta_0) + 2 \sum_{m=1}^{\infty} J_{2m}(\delta_0) \sin(2m\omega t) \\ &= J_0(\delta_0) + 2J_2(\delta_0) \sin(2\omega t) \end{aligned} \quad (\text{I.34})$$

Seules les harmoniques d'ordre $m = 0, 1, 2$ sont retenues dans les expressions I.34, les ordres supérieurs ne contenant pas d'informations supplémentaires.

En pratique, le signal électrique détecté se met sous la forme suivante :

$$S(t) = S_0 + S_1 \sin \omega t + S_2 \sin 2\omega t + \dots \quad (\text{I.35})$$

Les quantités mesurées par détection synchrone sont donc S_1 et S_2 . En remplaçant $\sin \delta(t)$ et $\cos \delta(t)$ par leurs expressions I.34 dans l'équation I.29, nous pouvons déduire les valeurs de I_s et I_c par identification avec l'expression I.35.

Remarque : Notons aussi que dans le cas de la mesure en transmission, les expressions I.30, I.31 et I.32 restent valables si l'on remplace r_p et r_s par respectivement t_p et t_s .

L'utilisation de l'ellipsomètre nécessite souvent le réglage des angles θ_P , θ_M et θ_A selon une disposition donnée. Deux configurations sont souvent utilisées dans la pratique. La première, notée *configuration II* par l'équipementier, est telle que $\theta_M = 0^\circ$ et $\theta_P = 45^\circ$ ($\theta_A - \theta_M = 45^\circ$). Cette configuration permet d'exprimer simplement I_s et I_c en fonction de ψ et Δ à partir des équations I.31 et I.32 par :

$$\begin{aligned} I_s &= \sin 2\psi \sin \Delta \\ I_c &= \sin 2\psi \cos \Delta \end{aligned} \quad (\text{I.36})$$

Dans la deuxième configuration, notée *configuration III*, on fixe $M = 45^\circ$ et $P = 45^\circ$ ($A - M = 45^\circ$). Les expressions de I_s et I_c deviennent :

$$\begin{aligned} I_s &= \sin 2\psi \sin \Delta \\ I_c &= \cos 2\psi \end{aligned} \quad (\text{I.37})$$

Notons que les intensités I_s et I_c varient entre les valeurs -1 et 1. L'utilisation de l'une ou de l'autre de ces deux configurations comporte des ambiguïtés sur la détermination de ψ et Δ pour certaines valeurs, comme on peut le voir mathématiquement dans les équations I.36 et I.37. En effet, dans la configuration *II* on remarque que la valeur de ψ est indéterminée entre ψ et $90^\circ - \psi$ par la seule connaissance de $\sin 2\psi$. Il n'est donc pas possible de déterminer la valeur de 2ψ à partir des seules équations I.36. Néanmoins, la détermination de Δ selon cette configuration est exempte d'indétermination. Par contre, si on choisit de travailler avec la configuration *III*, la valeur de Δ est cette fois-ci indéterminée entre deux solutions possibles, Δ et $180^\circ - \Delta$. En effet, les équations I.37 ne permettent que la détermination de $\sin \Delta$ alors que ψ est déterminé sans ambiguïté.

Pour éviter ces désagréments, il faut déterminer Δ dans la configuration *II* et ψ dans la configuration *III*. En pratique, on choisit d'exploiter les spectres bruts de I_s et I_c exempts d'indétermination dans une configuration. Dans la suite, la configuration *II* sera choisie.

Remarque : Dans cette partie, nous avons présenté l'ellipsométrie classique. Dans ce cas, le matériau est supposé homogène et isotrope. Les éléments hors diagonaux de la matrice de Jones de l'échantillon sont nuls (expression I.27). Il est tout de même possible d'utiliser l'ellipsométrie pour mesurer des matériaux possédant une anisotropie optique. On parle dans ce cas de l'ellipsométrie généralisée. La différence avec le cas classique est la représentation matricielle de l'échantillon qui devient :

$$S = \begin{bmatrix} r_{pp} & r_{sp} \\ r_{ps} & r_{ss} \end{bmatrix} \quad (\text{I.38})$$

Dans ce manuscrit, nous ne traiterons pas ce cas puisqu'on mesure pas les éléments non diagonaux r_{sp} et r_{ps} .

Nous venons d'exposer les fondements théoriques de l'ellipsométrie dans le cas classique. Le principe de mesure des angles ellipsométriques a été également détaillé, en insistant sur les techniques usuelles, à savoir l'ellipsométrie à annulation et l'ellipsométrie à modulation. Nous avons discuté les problématiques relatives à chacune de ces techniques. La partie suivante porte essentiellement sur la description de structures typiques qui seront caractérisées par la suite.

I.2 Principales structures caractérisables

Nous commencerons cette partie par aborder le cas le plus général d'une structure multicouche où plusieurs couches de matériaux peuvent être déposées sur un substrat. Dans le cadre particulier de notre travail, nous nous sommes volontairement limités au cas d'une seule couche déposée sur un substrat afin de valider notre méthode. Ensuite,

nous traiterons le cas d'une structure plus complexe présentant une modulation périodique de la surface : le réseau de diffraction, qui constitue la structure de référence pour étudier la caractérisation de nanostructures.

Pour toutes ces structures, nous présenterons succinctement le principe du calcul théorique des coefficients ellipsométriques ψ et Δ conduisant à la détermination des intensités I_s et I_c (définies par les équations I.36 et I.37) réellement utilisées par la suite. Pour tous les échantillons caractérisés dans ce manuscrit, les propriétés optiques des substrats ont été soit mesurées par ellipsométrie au laboratoire soit fournis par les différents partenaires. Dans tout ce qui va suivre, nous ne prendrons pas en compte d'éventuelles rugosités de surface.

Remarque : Notons que la structure la plus simple à caractériser est sans contexte un substrat massif d'un matériau donné. La représentation schématique de cette structure est identique à celle de la figure I.2. Les coefficients de réflexion r_p et r_s d'un substrat sont équivalents à ceux relatif à un simple interface séparant deux milieux d'indices différents (équation I.6 et I.7) si l'épaisseur du matériau est suffisamment grande pour considérer un milieu semi-infini.

I.2.1 Structure multicouche isotrope

I.2.1.1 Présentation et applications

La structure multicouche est constituée d'un empilement de m couches de matériaux différents supposés homogènes et isotropes, intercalées entre le milieu extérieur 0 et le substrat (milieu $m+1$) comme le montre la figure I.6. L'indice de la couche i est noté N_i et d_i son épaisseur.

Une onde incidente à la surface supérieure de la structure est, d'une part réfléchiée dans le même milieu (d'indice N_0), et d'autre part transmise à travers le substrat (d'indice N_{m+1}) après traversée des m couches (d'indices respectifs N_m).

De nombreuses applications en électronique et en optoélectronique utilisent des matériaux sous forme de couches minces (ou multicouches) déposées sur un substrat.

Avec les progrès réalisés dans la fabrication des transistors à couches minces [32] (en anglais TFT pour Thin-Film Transistor), ces structures ont reçu une attention considérable notamment dans la conception des écrans à cristaux liquides [33], ainsi que pour les systèmes de détection en biologie et en chimie [34].

La fabrication de microstructures (en optique ou en microélectronique) [7] par photolithographie nécessite généralement le dépôt d'une couche mince en résine, qui servira de masque après sa structuration. Il est ainsi souhaitable de déterminer précisément ses propriétés, optiques et géométriques, afin de bien maîtriser les conditions de gravure.

I.2.1.2 Formalisme matriciel

Le calcul théorique des angles ellipsométriques ψ et Δ est plus compliqué que le cas d'une simple interface (figure I.2), puisqu'il nécessite le calcul global des coefficients r_p et r_s qu'on détaillera par la suite.

Pour le traitement d'une telle structure, on emploiera plus facilement une approche matricielle [18] basée sur le fait que la propagation de la lumière est linéaire et sur les

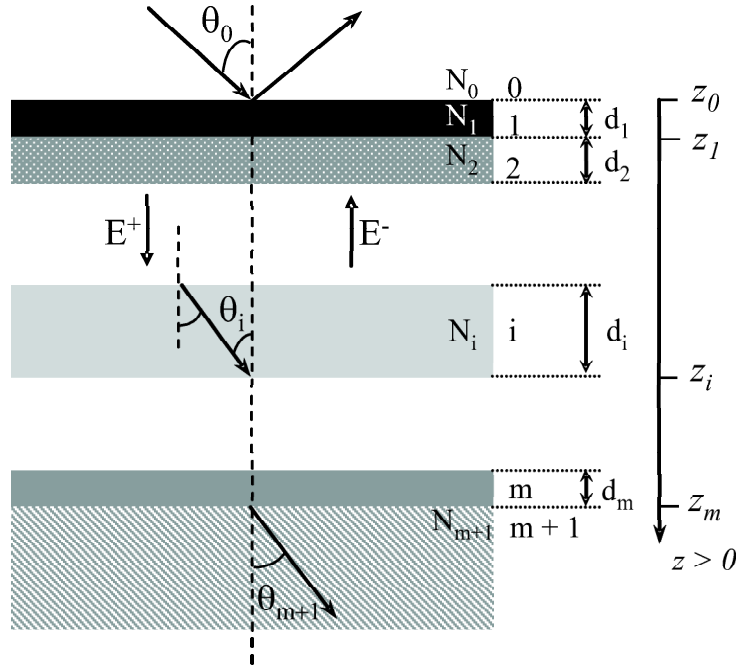


FIG. I.6: Propagation d'une onde plane à travers une structure multicouche constituée de m couches d'épaisseurs respectifs d_m . θ_0 est l'angle d'incidence dans le milieu 0 ; θ_i et θ_{m+1} sont les angles de réfraction respectivement dans la couche i et dans le substrat. On note E^+ et E^- les amplitudes des ondes se propageant respectivement dans le sens positif (allant du milieu 0 vers le substrat) et négatif. On désigne par z_i l'abscisse de chaque interface séparant le milieu i et $i+1$.

propriétés de continuité du champ à travers chaque interface séparant deux milieux isotropes.

L'amplitude de l'onde se propageant dans le sens positif (suivant Oz) est notée E^+ , et celle se propageant dans le sens négatif est notée E^- .

Le champ total à l'intérieur de la couche i , qui est excité par l'onde incidente, est composé de deux ondes planes : une se propageant dans le sens (+) ou $z > 0$ et l'autre dans le sens (-) ou $z < 0$. Les vecteurs d'onde correspondant sont contenus dans le même plan d'incidence.

A l'abscisse z , le champ total peut ainsi être représenté par le vecteur :

$$E(z) = \begin{bmatrix} E_z^+ \\ E_z^- \end{bmatrix} \quad (\text{I.39})$$

Les champs électriques dans les deux milieux externes 0 et $m+1$ doivent être liés entre eux par une matrice de transfert 2×2 notée S et telle que :

$$\begin{bmatrix} E^+(z_0^-) \\ E^-(z_0^-) \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} E^+(z_m^+) \\ E^-(z_m^+) \end{bmatrix} = [S] \begin{bmatrix} E^+(z_m^+) \\ E^-(z_m^+) \end{bmatrix} \quad (\text{I.40})$$

avec

- z_0^- : point d'abscisse z_0 dans la couche 0 à l'interface avec la couche 1,
- z_m^+ : point d'abscisse z_m dans la couche $m+1$ à l'interface avec la couche m .

S est la matrice caractérisant la structure multicouche confinée entre les deux milieux extrêmes. Elle peut être exprimée comme un produit de matrices *interface* $I_{i,i+1}$, carac-

térisant l'interface entre les couches i et $i + 1$ et de matrices *couche* (ou *propagation*) L_i , caractérisant la propagation de la lumière dans la couche i . Ce produit matriciel, noté S , est écrit de gauche à droite en suivant le parcours de l'onde dans la structure jusqu'au substrat :

$$S = I_{0,1} L_1 I_{1,2} \dots I_{i-1,i} L_i \dots L_m I_{m,m+1} \quad (\text{I.41})$$

La continuité des champs à l'interface des deux couches i et $i + 1$ s'écrit :

$$\begin{bmatrix} E^+(z_i^-) \\ E^-(z_i^-) \end{bmatrix} = I_{i,i+1} \begin{bmatrix} E^+(z_i^+) \\ E^-(z_i^+) \end{bmatrix} \quad (\text{I.42})$$

avec

$$I_{i,i+1} = \frac{1}{t_{i,i+1}} \begin{bmatrix} 1 & r_{i,i+1} \\ r_{i,i+1} & 1 \end{bmatrix} \quad (\text{I.43})$$

$r_{i,i+1}$ et $t_{i,i+1}$ étant les coefficients de réflexion et de transmission à l'interface des couches i et $i + 1$ pour une polarisation donnée, déjà exprimés dans les équations I.6 et I.7.

La propagation de l'onde à travers la couche i sera représentée simplement par :

$$\begin{bmatrix} E^+(z_{i-1}^+) \\ E^-(z_{i-1}^+) \end{bmatrix} = L_i \begin{bmatrix} E^+(z_i^-) \\ E^-(z_i^-) \end{bmatrix} \quad (\text{I.44})$$

avec

$$L_i = \begin{bmatrix} \exp(j\beta_i) & 0 \\ 0 & \exp(-j\beta_i) \end{bmatrix} \quad (\text{I.45})$$

avec $\beta_i = (2\pi/\lambda)d_i N_i \cos \theta_i$ est le déphasage introduit lorsque l'onde traverse la couche i , et θ_i , l'angle d'incidence sur la couche i , tel que $N_i \sin \theta_i = N_{i+1} \sin \theta_{i+1}$.

Ainsi, la matrice S est complètement définie en reportant I.43 et I.45 dans I.41. L'expression des amplitudes des champs à la surface prend la forme :

$$\begin{bmatrix} E^+(z_0^-) \\ E^-(z_0^-) \end{bmatrix} = S \begin{bmatrix} E^+(z_m^+) \\ 0 \end{bmatrix} \quad (\text{I.46})$$

sachant que l'amplitude du champ électrique se propageant dans le sens négatif dans le substrat $E^-(z_m^+)$ est nulle dans le cas où seul le superstrat est éclairé.

Finalement, les coefficients de réflexion et de transmission résultants dans une polarisation donnée nécessaires pour le calcul des angles ellipsométriques ψ et Δ d'une telle structure, peuvent être déduites de l'équation I.40 par :

$$r = \frac{E^-(z_0^-)}{E^+(z_0^-)} = \frac{S_{21}}{S_{11}} \quad (\text{I.47})$$

$$t = \frac{E^+(z_m^+)}{E^+(z_0^-)} = \frac{1}{S_{11}} \quad (\text{I.48})$$

En supposant que S^p et S^s représentent les matrices de transfert correspondant aux

cas des deux polarisations TM et TE , alors :

$$r_p = \frac{S_{21}^p}{S_{11}^p} ; \quad r_s = \frac{S_{21}^s}{S_{11}^s} \quad (I.49)$$

$$t_p = \frac{1}{S_{11}^p} ; \quad t_s = \frac{1}{S_{11}^s} \quad (I.50)$$

Remarque : Notons que S^p et S^s ne sont pas identiques à cause de la différence des matrices *interface* (I.43) qui apparaissent dans l'équation I.46 faisant intervenir les coefficients de réflexion et de transmission à chaque interface défini par les équations I.6-I.9. Par contre, les matrices *propagation* (I.45) restent les mêmes quel que soit la polarisation.

Un exemple simple de structure multicouche utilisant ce formalisme est celui d'un film d'épaisseur d , d'indice N_1 déposé sur un substrat, comme le montre la figure I.7.

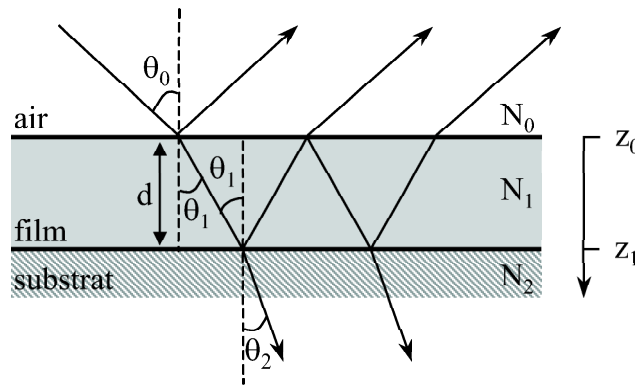


FIG. I.7: Réflexion et transmission d'une onde plane à travers une couche d'épaisseur d et d'indice N_1 déposée sur un substrat d'indice N_2 . L'angle d'incidence sur l'interface du film est noté θ_0 . Les angles de réfraction dans le film et le substrat sont notés θ_1 et θ_2 .

La matrice de transfert à travers cette structure est donnée par :

$$S = I_{01} L_1 I_{12} \quad (I.51)$$

En substituant les matrices I et L par leurs expressions I.43 et I.45, cette relation devient :

$$\begin{aligned} S &= \left(\frac{1}{t_{01}t_{12}} \right) \begin{bmatrix} 1 & r_{01} \\ r_{01} & 1 \end{bmatrix} \begin{bmatrix} \exp(j\beta) & 0 \\ 0 & \exp(-j\beta) \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \\ &= \left(\frac{\exp(j\beta)}{t_{01}t_{12}} \right) \begin{bmatrix} (1 + r_{01}r_{12} \exp(-2j\beta)) & (r_{12} + r_{01} \exp(-2j\beta)) \\ (r_{01} + r_{12} \exp(-2j\beta)) & (r_{01}r_{12} + \exp(-2j\beta)) \end{bmatrix} \end{aligned} \quad (I.52)$$

avec $\beta = (2\pi/\lambda)dN_1 \cos\theta_1$.

Il est alors facile de déterminer les coefficients de réflexion et de transmission par les équations I.47 et I.48 :

$$r = \frac{S_{21}}{S_{11}} = \frac{r_{01} + r_{12} \exp(-2j\beta)}{1 + r_{01}r_{12} \exp(-2j\beta)} \quad (I.53)$$

$$t = \frac{1}{S_{11}} = \frac{t_{01}t_{12} \exp(-j\beta)}{1 + r_{01}r_{12} \exp(-2j\beta)} \quad (I.54)$$

Connaissant l'indice de réfraction et l'épaisseur d de la couche, les angles ellipsométriques ψ et Δ peuvent être facilement calculés par l'équation I.14, en explicitant les coefficients I.53 et I.54 pour les deux polarisations TE et TM . Les intensités I_s et I_c correspondantes sont ensuite déterminées par les équations I.36 et I.37.

Remarque : Dans le cas d'une seule couche déposée sur le substrat, ce calcul peut être mené par l'addition des réflexions multiples à l'intérieur de la couche (figure I.7). Par contre, cette technique devient fastidieuse lorsque le nombre de couches devient supérieur à 2.

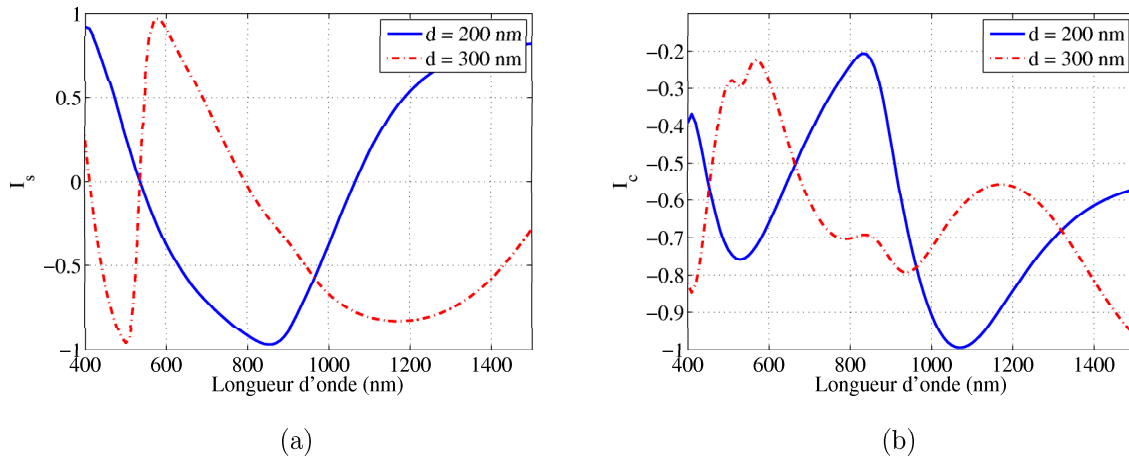


FIG. I.8: Intensités I_s (a) et I_c (b) calculées dans la gamme du visible au proche infrarouge, pour deux échantillons théoriques constitués respectivement d'une couche de résine d'épaisseur 200 nm et 300 nm, déposée sur un substrat de silicium.

Les figures I.8(a) et I.8(b) montrent les intensités I_s et I_c calculées à partir de ψ et Δ en configuration II, dans la gamme spectrale du visible au proche infrarouge pour une couche mince de résine photosensible déposée sur un substrat de silicium. Les indices optiques des matériaux sont connus (donnés dans l'annexe, la résine étant notée RES 1), et l'angle d'incidence θ_0 est fixé à 60° . Deux échantillons ayant une épaisseur de couche de résine différente (200 nm et 300 nm) sont représentés.

Ce type de calcul fait partie intégrante du logiciel d'exploitation de l'ellipsomètre. Toutefois, pour des raisons de compatibilité et de rapidité, il a fallu le reprogrammer dans le langage utilisé pour les réseaux de neurones.

Le deuxième volet de cette partie traitera des structures plus complexes qui sont modulées périodiquement en surface.

I.2.2 Structure périodique diffractante

I.2.2.1 Présentation et applications

La principale famille de structures abordée dans ce travail, est celle des structures périodiques. La structure de base qui sera traitée dans ce manuscrit est le réseau de diffraction. Nous nous intéresserons plus particulièrement au réseau dont la surface est modulée périodiquement, appelé aussi *réseau de surface* (en anglais *corrugated grating*).

En raison de leur périodicité, ces structures sont diffractantes. L'une des caractéristiques principales d'un réseau est sa période, notée Λ . Elle définit l'application spécifique selon son ordre de grandeur. On distingue deux géométries souvent employées dans la littérature : la première, notée $1D$, où la structure est périodique selon une seule direction (figure I.9(a)) ; et la deuxième, notée $2D$, où la structure est périodique selon deux directions orthogonales (figure I.9(b)). Le réseau $1D$ est formé de traits parallèles appelés aussi *sillons*. Ces derniers peuvent avoir n'importe quel forme de profil géométrique (lamellaire, trapézoïdal, sinusoïdal, etc. . .) à condition que la structure reste périodique. Les réseaux $2D$ sont formés de plots qui peuvent également avoir différentes formes à condition de garder la périodicité dans les deux directions.

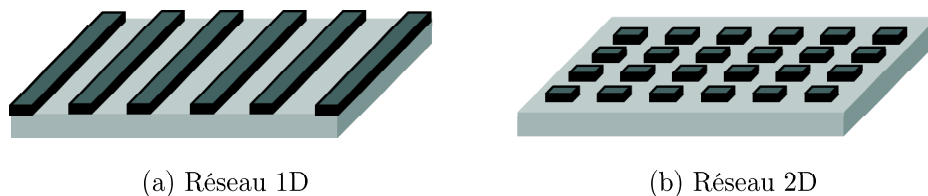


FIG. I.9: Modèles de réseau de diffraction lamellaire(a) $1D$, (b) $2D$.

Remarque : Dans la littérature scientifique, certains auteurs préfèrent définir un réseau par sa fréquence, désignant ainsi le nombre de traits par millimètre. Par exemple, un réseau de période $0,5 \mu\text{m}$ comporte 2000 traits/mm.

Nous limiterons volontairement notre étude au cas des réseaux $1D$ pour valider notre méthode de caractérisation. Toutefois, la démarche est transposable au cas des réseaux $2D$ dont la caractérisation présente aujourd'hui un enjeu majeur [35].

Les réseaux de diffraction possèdent une grande importance en spectroscopie, spécialement pour leur capacité à séparer ou disperser une lumière incidente dans plusieurs directions privilégiées. Cette propriété intéressante est largement mise à profit dans de nombreuses applications, dans des domaines aussi variés que l'optique, la photonique et l'électronique. En effet, le progrès technologique a permis la fabrication des réseaux sub-microniques pouvant ainsi être intégrés au sein même d'un dispositif. En parallèle, le développement des méthodes de calcul électromagnétique a aussi joué un rôle important dans la conception de ces composants notamment pour l'optimisation des paramètres de la structure selon l'application souhaitée.

Parmi les nombreux dispositifs utilisant seulement un réseau de diffraction, on peut citer les polariseurs fonctionnant en réflexion [36] ou en transmission [37, 38].

En optique guidée, l'intérêt du réseau réside dans l'efficacité du couplage de la lumière à l'intérieur des guides d'onde [39]. Dans le même domaine, il a été montré qu'en associant

un réseau de diffraction à une guide d'onde planaire, on peut réaliser des systèmes de biodétection [40].

Dans le domaine des lasers, la compression des impulsions femtosecondes est possible grâce aux réseaux résonnants. Ces réseaux peuvent être utilisés comme filtres à multiplexage par division de longueur d'onde [41]. Dans les diodes laser à cavité verticale émettant par la surface (en anglais VCSEL pour Vertical-Cavity Surface-Emitting Lasers), la polarisation de la lumière émise est mal définie et instable. Le réseau de diffraction, gravé à la surface, a permis de contrôler cet état de polarisation [42].

Le réseau de diffraction peut être employé dans les cellules solaires organiques [43] pour augmenter l'absorption du composant dans un certain domaine du spectre solaire.

Finalement, on doit noter que le réseau de diffraction est utilisé comme une structure test dans l'industrie de la microélectronique pour la fabrication de composants aux dimensions de plus en plus réduites. L'étude de cette structure n'est également pas anodine car elle sert de référence pour tester et comparer les différentes méthodes scatterométriques.

I.2.2.2 Phénomène de diffraction par un réseau

Considérons une onde plane de longueur d'onde λ_0 , de vecteur d'onde \vec{k}_i ($k_i = N_r k_0$, $k_0 = 2\pi/\lambda_0$) incidente sous un angle θ_i sur un réseau de période Λ séparant deux milieux d'indices respectifs N_r et N_t . L'angle d'incidence θ_i est défini par l'angle que fait la direction de propagation \vec{k}_i de l'onde incidente avec la normale au réseau (figure I.10). Définissons le vecteur réseau \vec{K}_r ($K_r = 2\pi/\Lambda$) par le vecteur contenu dans le plan du réseau et perpendiculaire aux sillons. Le plan d'incidence est le plan formé par le vecteur d'onde \vec{k}_i avec la normale au réseau. La direction angulaire θ_{dm} de l'onde diffractée dans l'ordre m est donnée par l'équation fondamentale des réseaux :

$$\begin{aligned} N_d \sin \theta_{dm} &= N_r \sin \theta_i + m \frac{K_r}{k_0} \\ &= N_r \sin \theta_i + m \frac{\lambda_0}{\Lambda}, \quad m = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (\text{I.55})$$

avec N_d : indice du milieu dans lequel on observe la diffraction, à savoir $-N_r$ pour les ordres diffractés en réflexion et N_t pour les ordres transmis à travers le réseau.

Par convention, on affecte le signe négatif aux ordres qui se diffractent à gauche de l'ordre zéro (transmis ou réfléchi), et positifs ceux qui se diffractent dans l'autre sens.

Remarque : Dans le cas particulier où $m = 0$, on retrouve les équations de Descartes (I.1 et I.2) décrivant le comportement de la lumière en transmission et réflexion à l'interface de deux milieux semi-infinis.

Le réseau se comporte ainsi comme un élément dispersif lorsque la lumière incidente est polychromatique. En effet, chaque angle de diffraction θ_{dm} ($m \neq 0$) dépend de la longueur d'onde λ_0 considérée (I.55), on peut ainsi séparer angulairement les différentes longueurs d'onde composant le spectre.

Les ordres diffractés dans la figure I.10 sont représentés dans l'espace direct, défini par le repère Oxz qui contient le plan d'incidence. Il est également possible de représenter les ordres de diffraction dans l'espace réciproque des fréquences spatiales dit *espace des "k"*. C'est de cette façon que l'on peut déterminer géométriquement les directions diffractées. Les composantes $(\vec{k}_{dm})_x$, sur l'axe x du réseau, des vecteurs d'ondes des différents ordres

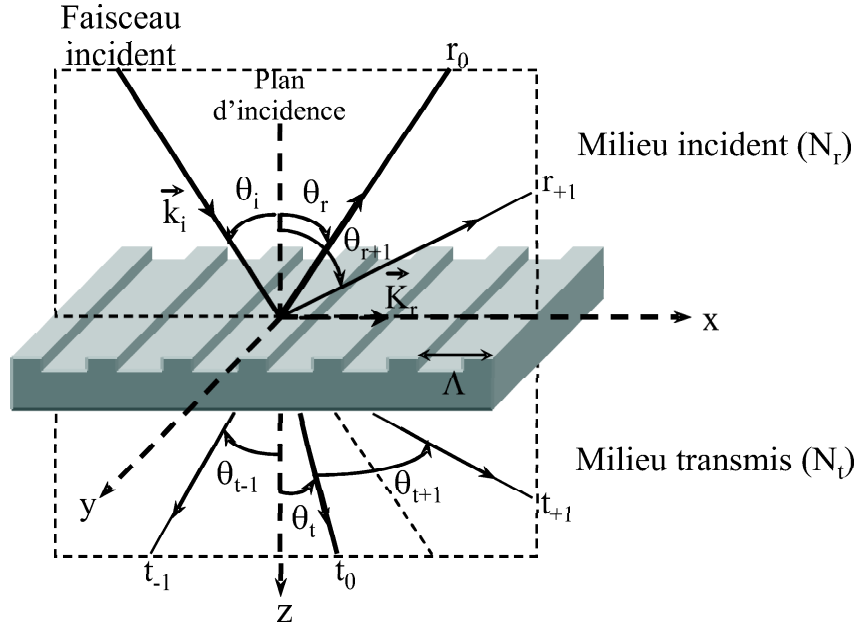


FIG. I.10: Représentation du phénomène de diffraction par un réseau de période Λ dans l'espace direct en configuration classique.

m diffractés sont exprimées en fonction de celles du vecteur d'onde incident $(\vec{k}_{i0})_x$ et du vecteur réseau par l'équation suivante déduite de l'équation I.55 :

$$(\vec{k}_{dm})_x = (\vec{k}_{i0})_x + m\vec{K}_r \quad \text{avec } m = 0, \pm 1, \pm 2, \dots \quad (\text{I.56})$$

Il est ainsi facile de déterminer l'orientation des différents ordres diffractés à partir de cette équation par simple construction géométrique (figure I.11). Tous les vecteurs d'ondes sont reportés dans le plan d'incidence défini par les axes Ok_x et Ok_y , et sont tracés à partir d'une origine commune O qui est le point d'impact du faisceau incident sur la surface du réseau. Ils sont compris dans les demi-cercles de rayon $N_r k_0$ pour le milieu réfléchi et $N_t k_0$ pour le milieu transmis. Pour construire ces vecteurs, on représente d'abord les vecteurs d'onde incident, réfléchi puis transmis d'ordre zéro en utilisant les lois de la réflexion et la réfraction de Descartes (I.1 et I.2). Les différents vecteurs d'ondes diffractés en réflexion (ou en transmission) sont ensuite obtenus en ajoutant ou en retranchant la quantité \vec{K}_r à la composante sur x du vecteur d'onde réfléchi (respectivement transmis) dans l'ordre zéro $(k_{r0})_x$ conformément à l'équation I.56. En diffraction classique, une onde incidente polarisée *TE* (respectivement *TM*) conserve son état de polarisation

Il faut cependant noter que tout ce qui précède concerne le cas d'une incidence classique où le plan d'incidence est perpendiculaire aux sillons, soit \vec{K}_r dans le plan d'incidence. Dans ce cas, on parle de la *diffraction classique* et tous les ordres diffractés, réfléchis ou transmis, sont contenus dans le plan d'incidence.

Il existe une autre configuration pour laquelle le plan d'incidence n'est pas perpendiculaire aux sillons du réseau, mais fait un angle φ (appelé aussi *azimut*). Dans ce cas, on parle de *diffraction conique* (figure I.12). Les ordres diffractés ne sont plus contenus dans un plan mais sont situés sur la surface de demi-sphères ; l'une en réflexion et l'autre en transmission de rayons respectives $N_r k_0$ et $N_t k_0$. Si le plan d'incidence est parallèle aux

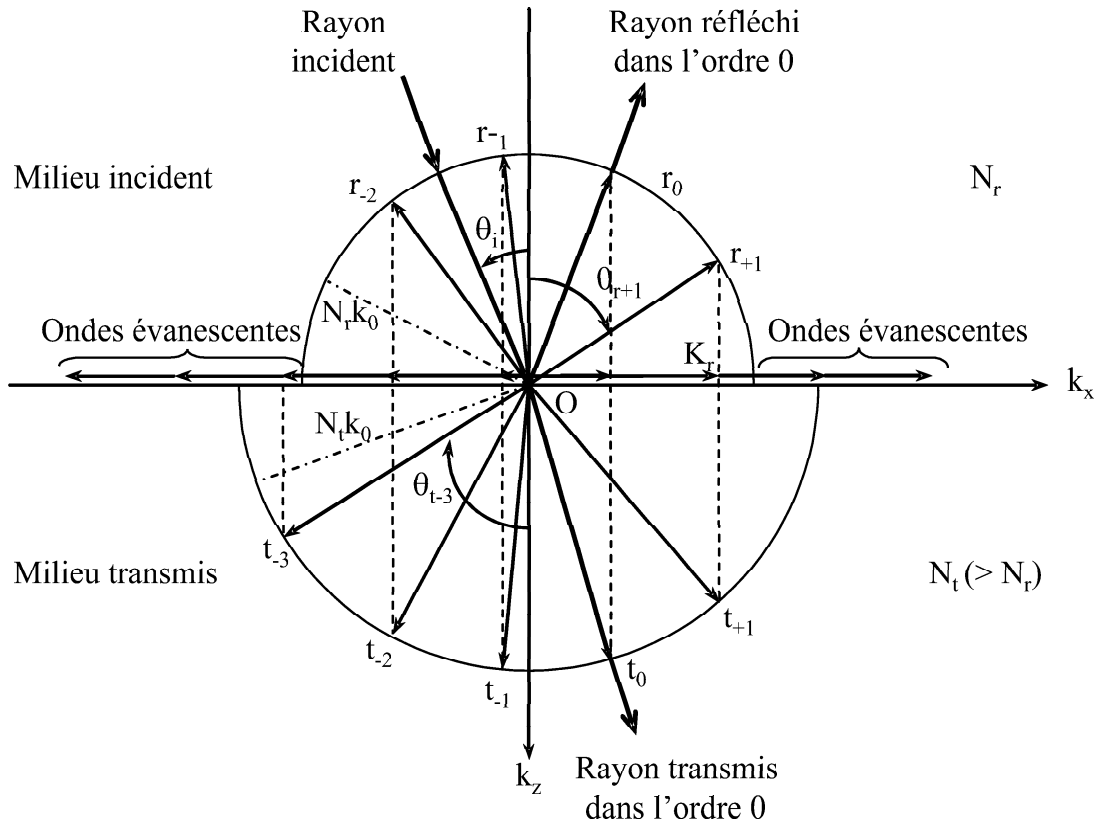


FIG. I.11: Représentation dans l'espace réciproque de la diffraction par un réseau de période Λ .

sillons ($\varphi = 90^\circ$), on parle de *diffraction pure conique*. La construction géométrique dans l'espace direct est similaire que le cas précédent en prenant en compte la composante suivant l'axe Oy. Il faut également noter l'apparition de conversion d'état de polarisation. Ainsi, une onde polarisée *TE* (respectivement *TM*) ne reste pas exclusivement *TE* (respectivement *TM*) après diffraction en configuration conique.

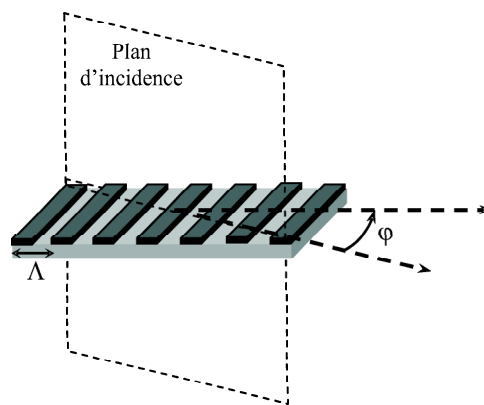


FIG. I.12: Représentation du phénomène de diffraction par un réseau de période Λ dans l'espace direct en configuration conique.

Il est intéressant d'examiner les conditions pour lesquelles les ordres peuvent exister pour un réseau donné. L'équation I.55 des réseaux ne possède de solutions que lorsque la condition suivante est vérifiée :

$$|\sin \theta_{dm}| < 1 \quad (\text{I.57})$$

Les ordres m sont alors des ordres propagatifs facilement détectables loin de la surface de diffraction. Dans le cas où cette inégalité n'est pas respectée, il existe cependant des ordres mais qui correspondent à des ondes dites *évanescentes* (figure I.10). Ces ondes sont plus difficilement mesurables et ne seront pas exploitées dans la suite de ce travail. Cependant, il faut noter que ces ondes doivent être prises en compte dans le calcul théorique de ψ et Δ . Ces ordres jouent un rôle important dans différentes applications des guides d'onde et des fibres optiques [44].

A titre d'exemple le tableau suivant donne le nombre d'ordres de diffraction pouvant se propager (en réflexion) en fonction de la période du réseau et de la longueur d'onde incidente λ_0 pour un angle θ_i fixé à 10° .

$\Lambda_0(nm) \backslash \lambda(nm)$	1000	800	400	200
2000	4	6	10	20
1000	2	3	5	10
500	1	1	3	5
200	1	1	1	2

TAB. I.1: Nombre d'ordres total (positifs et négatifs) diffractés sur une structure périodique en fonction de la période Λ (nm) et de la longueur d'onde λ_0 (nm) pour une onde incidente avec un angle $\theta_0 = 10^\circ$.

On remarque que plus le rapport λ/Λ est petit (c'est à dire plus la longueur d'onde est faible par rapport à la période du réseau), plus le nombre d'ordres diffractés est grand (équation I.55 et tableau I.1). Ceci peut facilement être expliqué par la figure I.11, le rayon $n_r k_0$ est plus grand si λ_0 diminue ce qui engendre par construction géométrique plus d'ordres. De la même façon, si la période Λ est grande, la quantité K_r est plus petite et le nombre d'ordres diffractés augmente. Dans le cas contraire, le nombre d'ordre diminue jusqu'à ce que seule la *réflexion spéculaire* soit observée (ordre 0).

Il existe une configuration particulière pour laquelle le faisceau incident coïncide avec un ordre m diffracté (en général l'ordre -1). On parle de configuration de *Littrow* ou d'*autocollimation*. L'utilité d'un tel montage n'est pas négligeable puisqu'il constitue une méthode simple et performante permettant de mesurer la période d'un réseau. En effet, pour un réseau de période Λ , la relation liant l'angle d'incidence dans la configuration de Littrow, noté θ_{li} , et la longueur d'onde λ_0 utilisée est obtenue à partir de l'équation I.55 en posant $\theta_{dm} = \theta_{li}$:

$$2n_d \sin \theta_{li} = m \frac{\lambda_0}{\Lambda} \quad (\text{I.58})$$

ce qui permet d'obtenir la valeur de la période λ :

$$\Lambda = \frac{m\lambda_0}{2n_d \sin \theta_{li}} \quad (\text{I.59})$$

Prenons l'exemple où $\lambda_0 = 400$ nm pour un réseau de période $\Lambda = 565$ nm. D'après l'équation I.58 ($n_d = 1$ et $m = -1$), l'angle de littrow correspondant est égale $\theta_{li} = 20,73^\circ$. En utilisant la loi de propagation des erreurs, l'incertitude obtenue sur la période est donnée par :

$$\begin{aligned}
\sigma_{\Lambda}^2 &= \left(\frac{\partial \Lambda}{\partial \lambda_0} \right)^2 \sigma_{\lambda_0}^2 + \left(\frac{\partial \Lambda}{\partial \theta_{li}} \right)^2 \sigma_{\theta_{li}}^2 \\
&= \left(\frac{m}{2n_d \sin \theta_{li}} \right)^2 \sigma_{\lambda_0}^2 + \left(\frac{m\lambda_0 \cos \theta_{li}}{2n_d \sin^2 \theta_{li}} \right)^2 \sigma_{\theta_{li}}^2
\end{aligned} \tag{I.60}$$

Si on suppose que λ_0 et θ_{li} sont connus avec des incertitudes respectives de l'ordre de $\sigma_{\lambda_0} = 0,5 \text{ nm}$ et $\sigma_{\theta_{li}} = 0,01^\circ$, on a ainsi $\sigma_{\Lambda} = 1,4 \text{ nm}$.

Cette méthode permet la détermination de la période avec une bonne précision. Toutefois, cette condition n'est pas toujours satisfaite lorsque le réseau sous test est de très faible période même pour des courtes longueurs d'onde. Dans ce cas, l'ordre -1 disparaît comme on vient de le voir dans le tableau I.1. La détermination de la période par la technique de Littrow n'est donc plus réalisable. Nous proposerons par la suite une méthode alternative pour la détermination de ce paramètre quelque soit son ordre de grandeur.

I.2.2.3 Calcul électromagnétique du champ diffracté

L'équation des réseaux (I.55) donne la direction des diverses ondes planes diffractées, réfléchies et transmises. Par contre, elle ne fournit aucune information concernant la distribution de l'énergie dans ces directions. Le calcul des intensités des champs diffractés par un réseau présente un degré de complexité supérieur à celui des structures multicouches, et ne possède pas jusqu'à présent de solution analytique.

Deux familles de méthodes existent pour achever ce type de calcul : les méthodes *approchées* et les méthodes *rigoureuses* qualifiées d'*exactes*. Les méthodes approchées nécessitent des hypothèses astreignantes, tandis que les méthodes rigoureuses consistent à résoudre les équations de Maxwell sans aucune approximation sur le plan théorique ; les seules approximations sont numériques (troncature du nombre d'équations, erreurs d'arrondis, ...).

Méthodes approchées. La *théorie scalaire* est la méthode la plus simple et la plus ancienne qui traite la propagation de la lumière comme un phénomène scalaire sans considérer sa nature électromagnétique. Elle permet de résoudre analytiquement le problème de la diffraction dans des cas limités, lorsque l'incidence θ_i est presque normale et les dimensions du réseau sont grandes par rapport à la longueur d'onde λ .

Cette théorie ne prenant pas en compte le caractère vectoriel du champ électromagnétique, le phénomène de diffraction est supposé indépendant de la polarisation. Elle ne peut pas donc expliquer les anomalies observées par Wood [45] (Wood avait mis en évidence des changements brusques de l'intensité de la lumière diffractée dans le cas des réseaux fortement conducteurs comme l'argent ou l'aluminium). Le seul avantage de cette méthode est sa simplicité et la facilité d'interpréter les résultats. Toutefois, elle reste très peu employée aujourd'hui.

Puis, ce fut Lord Rayleigh qui, pour la première fois, essaya de traiter le problème de diffraction par la voie électromagnétique. L'hypothèse de Rayleigh exprime le champ électromagnétique, à l'extérieur et à l'intérieur de la région modulée, par une série d'ondes planes appelée le *développement de Rayleigh*. Bien que ce développement reste toujours valide en dehors du réseau, il ne l'est pas à l'intérieur dans le cas où le réseau est profond. Il a été démontré [46] que pour un réseau sinusoïdal par exemple, cette théorie est bien

adaptée à condition que $h/\Lambda < 0.15$ (h étant la hauteur du réseau).

Méthodes rigoureuses. La *méthode intégrale* est l'une des premières méthodes rigoureuses [47, 48, 49] pour le calcul de la diffraction des réseaux. En utilisant le théorème de Green, le champ électrique peut être représenté en chaque point de l'espace par une intégrale sur la surface du réseau. L'onde incidente génère un courant de surface et le champ diffracté est obtenu par le rayonnement de ce courant tout le long de la surface du réseau. Ce courant dépend du champ incident ainsi que du champ diffracté en d'autres points de la surface. Il se calcule à partir d'une équation intégrale qui fait intervenir les conditions aux limites sur la surface du réseau. Cette méthode est performante et capable de traiter un très grand nombre de réseaux différents. Son inconvénient majeur provient de la complexité de la formulation mathématique et d'une certaine difficulté à la mettre en œuvre.

Il existe d'autres méthodes qui peuvent être qualifiées d'intégrales, comme la *méthode des éléments finis* [50] et la *méthode des sources fictives* [51].

Dans la première méthode, un maillage spécial est utilisé pour représenter les différents espaces d'étude. Le champ total est exprimé comme une somme de fonctions élémentaires sur toutes les cellules du maillage. Le problème est ainsi réduit à un système d'équations linéaires algébriques, avec comme inconnues les amplitudes des champs diffractés. Bien que l'idée de traitement soit simple, l'algorithme de maillage présente une complexité remarquable. Par exemple, dans le cas des réseaux fortement conducteurs le maillage doit tenir compte des changements rapides du champ à l'intérieur du métal. Le temps de calcul devient également très long.

Dans la deuxième méthode, le champ diffracté est représenté par un champ créé par des sources fictives (n'ayant pas d'existence physique). Ces dernières doivent être choisies de sorte que le champ total vérifie les conditions de passage associées aux équations de Maxwell. Par contre, il n'existe pas un algorithme général permettant de choisir la position et le type des sources, ce qui complique sa mise en œuvre pour les non spécialistes. Elle nécessite une certaine expérience préalable. La *méthode des sources généralisée* [52, 53] est une méthode connexe qui utilise un formalisme similaire.

La *méthode différentielle* est une méthode rigoureuse où les équations de Maxwell sont exprimées sous leur forme différentielle [54, 55]. Le champ électrique et la permittivité à l'intérieur du réseau sont représentés par une série de Fourier, et un système d'équations différentielles ordinaires est à résoudre après avoir retenu un nombre fini d'harmoniques (nombre de termes de la série de Fourier). Récemment cette méthode a été reformulée en y introduisant les règles de factorisation à partir des composantes normales et tangentielles des champs mais en conservant le système de coordonnées cartésien [56, 57]. L'algorithme de la matrice R [58] y a également été introduit. Ce dernier permet de relier aux mêmes points de l'espace les amplitudes du champ électrique à celles du champ magnétique. Aujourd'hui, il est admis que cette méthode est efficace en polarisation TM et pour les réseaux métalliques qui constituaient les deux principales limitations initiales. Elle est rapide et efficace sur une large gamme spectrale pour un très grand nombre de réseaux.

Il existe une autre approche qui peut être également qualifiée de méthode différentielle. Cette méthode est connue sous le nom de *méthode C*, où l'on effectue un changement de coordonnées simples [59, 60] pour redresser le profil et séparer l'espace en deux milieux

homogènes. Elle est valable pour les réseaux à profil continu et dérivable. Les réseaux purement lamellaires ne peuvent donc pas être traités. Elle fonctionne aussi bien en polarisation TE qu'en polarisation TM . Sa limitation est purement géométrique essentiellement liée à la pente du profil.

Enfin, la méthode qui suit est initialement adaptée aux réseaux lamellaires, mais est facilement étendue aux profils de forme plus complexe. En effet, la *méthode modale* exprime les solutions des équations de Maxwell sous la forme de modes de propagation dans le réseau [61, 62]. Le profil géométrique du réseau doit être approché par un empilement de couches en marches d'escalier, comme l'on montre sur la figure I.13. Le champ total est donné par une combinaison linéaire de tous les modes de chaque couche et les amplitudes sont déterminées par les conditions aux limites. Cette méthode est communément appelée MMMFE (de l'anglais Multilayer Modal Method by Fourier Expansion).

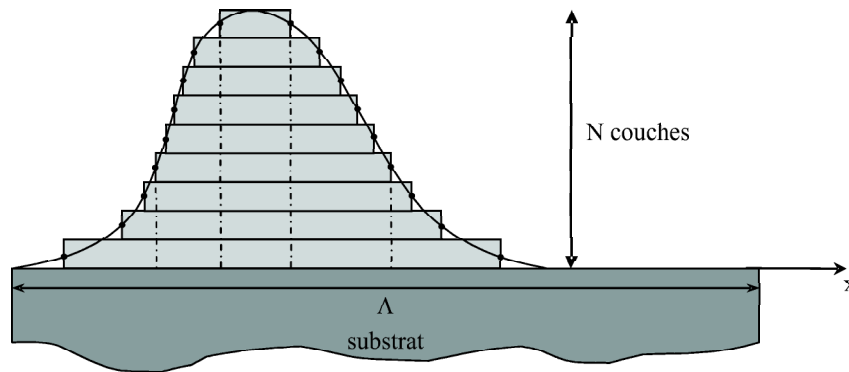


FIG. I.13: Découpage du profil géométrique quelconque d'un réseau en couches lamellaires.

L'origine de cette méthode remonte au début des années 80, lorsque Moharam et Gaylord en ont publié la première version connue à l'époque sous le nom *théorie rigoureuse des ondes couplées* (en anglais RCWA pour *Rigorous Coupled Wave Analysis*) [63].

A l'intérieur du réseau, le champ électrique est décomposé en série de Fourier généralisée en utilisant le théorème de Floquet (ou théorème de Bloch). A l'extérieur, on se sert du développement de Rayleigh pour exprimer les champs. On représente le réseau par sa permittivité $\epsilon(x)$ qui est une fonction périodique donnée par la décomposition suivante :

$$\epsilon(x) = \sum_{n=-m}^{+m} \epsilon_n \exp\left(j \frac{2\pi n x}{\Lambda}\right) \quad (\text{I.61})$$

où les ϵ_n sont les coefficients de Fourier.

Numériquement, un nombre limité d'harmoniques $M = 2m+1$ est retenu pour effectuer le calcul. La méthode se ramène à la résolution d'un système de N équations différentielles.

Cette méthode a été appliquée pour résoudre le problème de la diffraction dans le cas des réseaux de profils quelconques, absorbants et métalliques [64, 65, 66]. Elle s'est ensuite adaptée au cas de la diffraction conique [67] ainsi qu'aux réseaux à deux dimensions [68]. Elle a également traité les réseaux de diffraction anisotropes [69]. Suite à des limitations relatives à l'épaisseur de la structure, deux approches peuvent être utilisées : l'algorithme de la matrice R [58], et celui de la matrice S permettant de relier les ondes entrantes aux ondes sortantes pour une même couche de la structure [70]. Une reformulation a été ensuite

établie pour améliorer la convergence de la méthode dans le cas des réseaux métalliques pour la polarisation TM [71, 72]. C'est L. Li [56] qui a donné une base mathématique solide à ces études premières en introduisant les règles connues sous le nom de règles de factorisation.

Les principales limitations de cette méthode sont liées au fait que le profil réel doit être approché par un empilement de couche. Pour les réseaux fortement métalliques (Argent par exemple), ce découpage en tranches favorise la présence de sources fictives induisant des erreurs dans le calcul. Toutefois, cette méthode est largement utilisée pour le calcul des réseaux, notamment en scatterométrie.

C'est cette méthode que nous adopterons par la suite. Les codes de calcul utilisés dans cette thèse ont été développés par l'équipe "Electromagnétisme" du Laboratoire LASMEA (UMR 6602 - CNRS, Université Blaise Pascal - Clermont Ferrand).

Nous venons de présenter les structures diffractantes utilisées par la suite, à savoir les réseaux de diffraction, ainsi que les différentes méthodes numériques permettant de simuler la réponse électromagnétique de ces dernières. Ce calcul des intensités diffractées nous permet d'accéder facilement aux angles ellipsométriques ψ et Δ (ainsi qu'aux intensités I_s et I_c) en considérant chaque ordre de diffraction dans le milieu incident comme une simple réflexion sur une surface plane. Ainsi, les lois énoncées auparavant peuvent être appliquées.

Comme il a été mentionné, le calcul électromagnétique par la méthode MMMFE d'un réseau de profil quelconque nécessite sa discrétisation en profondeur en considérant un ensemble de N couches suffisant pour approcher le profil réel. D'un autre côté, la précision dépend également du nombre d'harmoniques M retenus dans le développement de Fourier. Le but de cette partie est d'étudier l'influence de ces deux paramètres sur la convergence de la méthode pour un profil de réseau donné.

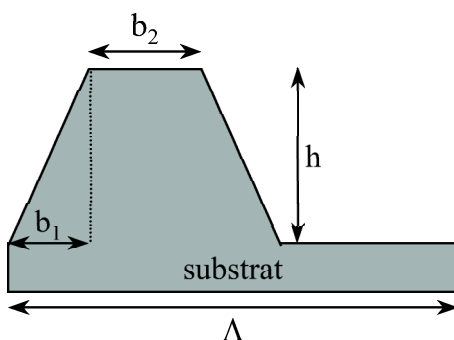


FIG. I.14: Schéma classique d'un réseau de diffraction de profil trapézoïdal symétrique défini par 3 paramètres : la projection de la pente b_1 , la largeur de la ligne b_2 et la hauteur h . La période Λ est supposée connue.

Considérons le cas d'un simple réseau de diffraction de profil trapézoïdal symétrique défini par trois paramètres géométriques : la projection de la pente b_1 , la largeur de la ligne b_2 (ou CD pour "critical dimension" en anglais) et la hauteur h (figure I.14), dont les valeurs sont respectivement 30 nm, 120 nm et 180 nm. La période est fixée à 565 nm. Les réseaux qui seront étudiés sont gravés dans deux types de matériaux, l'un absorbant (Silicium) et l'autre transparent (Silice).

En pratique, il est d'usage de définir le rapport ligne/espace $R_{l/e}$ (en anglais *line/space ratio*), le rapport d'aspect R_a (en anglais *aspect ratio*) et le rapport cyclique R_c (en anglais *duty cycle*). Ces définitions concernent les réseaux lamellaires ($b_1 = 0$), mais peuvent s'étendre aussi au cas du profil trapézoïdal. Le premier rapport caractérise la proportion de sillons par rapport aux espaces, le deuxième exprime l'importance de la profondeur par rapport aux sillons et le troisième correspond au rapport de la largeur du sillon sur la période. Ils sont donnés par :

$$R_{l/e} = \frac{b_2}{\Lambda - b_2} \quad (b_1 = 0) \quad \text{et} \quad R_{l/e} = \frac{b_2 + b_1}{\Lambda - (b_2 + b_1)} \quad (b_1 \neq 0) \quad (\text{I.62})$$

$$R_a = \frac{h}{\Lambda - b_2} \quad (b_1 = 0) \quad \text{et} \quad R_a = \frac{h}{\Lambda - (b_2 + b_1)} \quad (b_1 \neq 0) \quad (\text{I.63})$$

$$R_c = \frac{b_2}{\Lambda} \quad (b_1 = 0) \quad \text{et} \quad R_c = \frac{b_2 + b_1}{\Lambda} \quad (b_1 \neq 0) \quad (\text{I.64})$$

La signature ellipsométrique considérée est constituée de l'ensemble des intensités I_s et I_c pour l'ordre 0, calculée pour une plage de longueur d'onde comprise entre 350 nm et 700 nm par pas de 5 nm, dans le cas d'une incidence classique avec $\theta = 60^\circ$.

La convergence (C_M respectivement C_N) est évaluée pour la signature entière, comme étant la somme des dérivées de I_s et I_c calculées par rapport au paramètre étudié (M respectivement N). Elle est exprimée par les deux expressions suivantes, pour chacun des paramètres M et N :

$$C_M = \sum_{\lambda} \left(\left| \frac{\partial I_s}{\partial M} \right| + \left| \frac{\partial I_c}{\partial M} \right| \right) \quad (\text{I.65})$$

$$C_N = \sum_{\lambda} \left(\left| \frac{\partial I_s}{\partial N} \right| + \left| \frac{\partial I_c}{\partial N} \right| \right) \quad (\text{I.66})$$

Les figures I.15 représentent les variations de C_M en fonction du nombre de troncature M pour deux types de réseaux, l'un en silicium et l'autre en silice. Le nombre de couches N est fixé et supposé suffisant ($N = 20$) pour décrire le profil de la structure (figure I.14). Pour chacun des réseaux de périodes différentes, les rapports R_a et $R_{l/e}$ sont gardés constants, à savoir 0,43 et 0,36 respectivement.

Nous pouvons constater qu'au delà d'un certain nombre d'harmoniques supposé suffisant, la valeur calculée de I_s et I_c n'évolue pratiquement plus. Il faut également noter que pour un matériau donné, la convergence est plus rapide pour les réseaux à faibles périodes. Ceci est dû au fait que le nombre d'ordre de diffraction diminue quand le rapport λ/Λ augmente (tableau I.1). Ainsi, si le nombre de troncature M est insuffisant, les intensités diffractées dans chaque ordre sont calculées d'une façon approximative ce qui aboutit à une imprécision. D'autre part, la convergence est plus rapide, pour un même nombre M , pour un matériau transparent comme la silice que pour un matériau absorbant comme le silicium. Ceci provient probablement du fait que l'indice de réfraction est complexe.

Il est difficile d'interpréter quantitativement les valeurs de C_M et C_N . Ainsi, il nous est apparu plus parlant de calculer une erreur quadratique moyenne sur l'ensemble de la signature employée en fonction de M définie par :

$$E_{(I_s, I_c)}^M = \sqrt{\frac{1}{N} \sum_{\lambda} ([I_s(M) - I_s(M-1)]^2 + [I_c(M) - I_c(M-1)]^2)} \quad (\text{I.67})$$

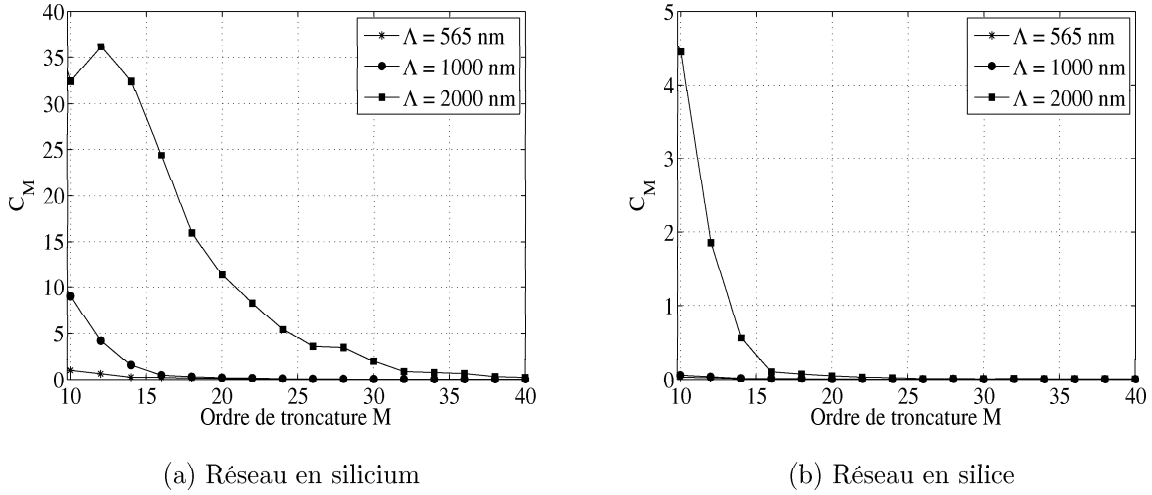


FIG. I.15: Convergence (C_M) de la MMMFE en fonction de l'ordre de troncature M , étudiée dans le cadre d'une signature ellipsométrique (I_s , I_c) sur deux réseaux de profil trapézoïdal pour différentes valeurs de la période Λ . R_a et $R_{l/e}$ sont maintenus fixe à respectivement 0,43 et 0,36. Le nombre de couches est fixé à 20.

Cette erreur sera comparée à l'erreur de mesure sur les intensités I_s et I_c de notre instrument à savoir de l'ordre de 10^{-2} . Prenons par exemple le cas du réseau de période 1000 nm ayant les paramètres $b_1 = 30$ nm, $b_2 = 200$ nm et $h = 260$ nm. L'erreur $E_{(I_s, I_c)}^M$ correspondante est représentée sur la figure I.16.

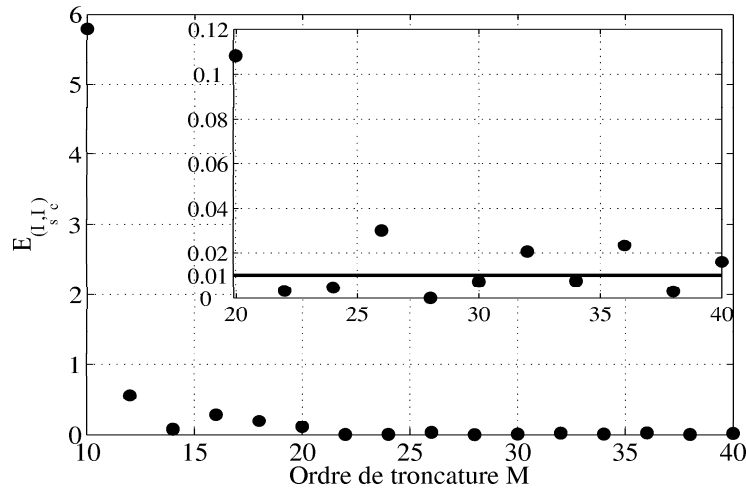


FIG. I.16: Variation de $E_{(I_s, I_c)}$ en fonction de M dans le cas du réseau de profil trapézoïdal en silicium ($b_1 = 30$ nm, $b_2 = 200$ nm, $h = 260$ nm) de période 1000 nm.

Suivant le critère fixé, nous pouvons constater que l'erreur $E_{(I_s, I_c)}$ se stabilise au delà de $M = 22$ avec des petites fluctuations autour de la valeur 10^{-2} (conformément aux spécifications du constructeur). Nous considérons donc que la convergence est ainsi atteinte pour cet ordre de troncature. Pour se rendre compte des éventuels biais introduits, nous avons reporté les signatures ellipsométriques calculées pour $M = 10$, 22 et 40 (figure I.17). La variation de I_s et I_c confirme le choix du nombre M assurant la convergence. Au delà de $M = 22$, les valeurs des intensités n'évoluent plus.

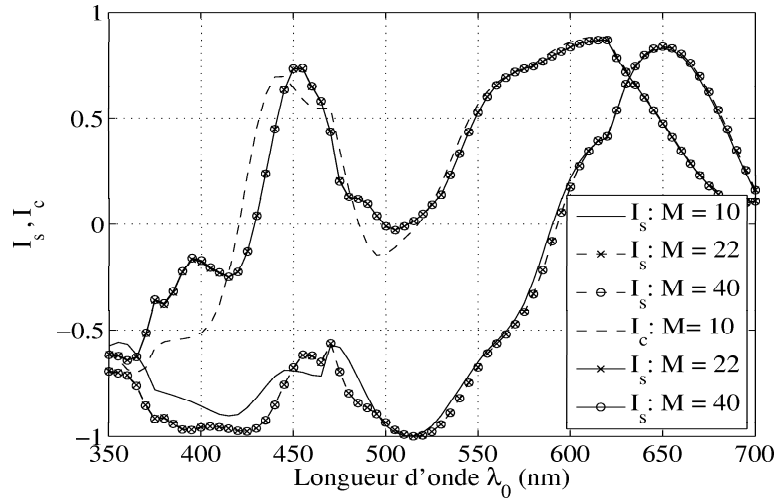


FIG. I.17: Signature ellipsométrique (I_s , I_c) calculée pour un réseau de profil trapézoïdal en silicium ($b_1 = 30$ nm, $b_2 = 200$ nm, $h = 260$ nm) de période 1000 nm, pour différents ordre de troncature.

Afin d'étudier à présent la convergence C_N en fonction du nombre de couches N , nous allons considérer deux réseaux de période $\Lambda = 565$ nm de profils différents : le premier P_1 est un trapèze simple, le deuxième P_2 possède des bords arrondis prononcés en haut et en bas. Les profils géométriques correspondants sur une période sont représentés sur la figure I.18. Notons que le nombre de troncature M est choisi assez large afin de minimiser l'influence de ce paramètre sur C_N .

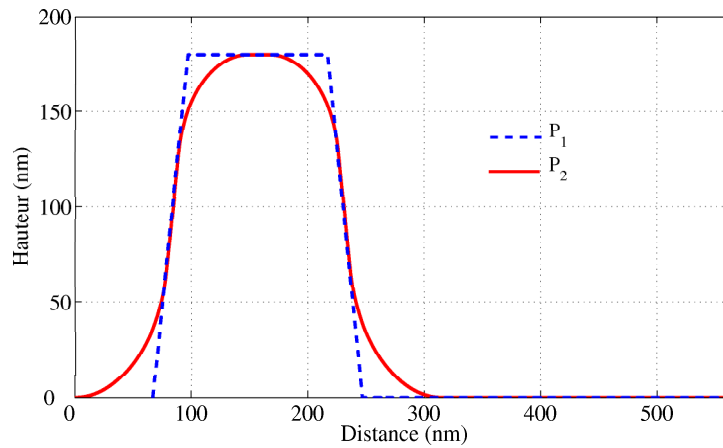


FIG. I.18: Représentation de deux profils de période $\Lambda = 565$ nm : trapézoïdal (P_1) et trapézoïdal présentant des arrondis en haut et en bas (P_2).

La convergence C_N (équation I.66) étudiée dans le cas des deux matériaux précédents (Silicium et Silice) est reportée dans la figure I.19 en fonction du nombre de couches. Le profil arrondi nécessite un nombre de couches plus élevé que le simple trapèze quelque soit le matériau.

En conclusion, le couple (M, N) qui garantit la convergence de la solution numérique varie selon le type de réseau étudié. Ainsi, ces deux paramètres doivent être judicieusement choisis en tenant compte de plusieurs paramètres, comme le matériau, la période, le profil du réseau, etc. . . .

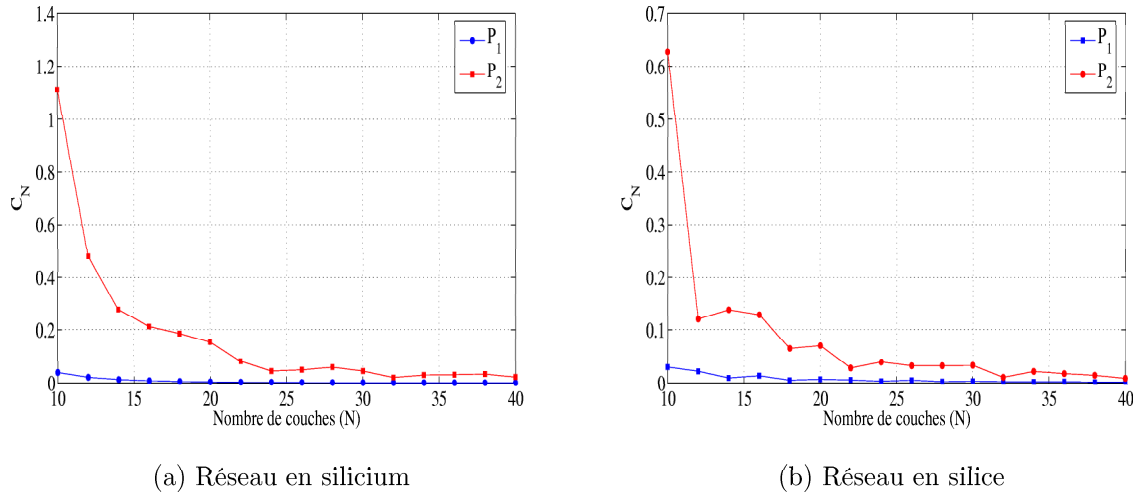


FIG. I.19: Convergence C_N de la méthode en fonction du nombre de couches N , étudiée dans le cadre d'une signature ellipsométrique (I_s , I_c) sur un réseau de profil trapézoïdal pour différentes valeurs de périodes.

Il serait ainsi tentant de prendre M et N assez grands dès le début afin d'éviter la recherche fastidieuse de leur valeur optimale pour chaque réseau étudié. Or, ce choix n'est pas raisonnable puisqu'il augmente considérablement le temps de calcul.

En effet, considérons le cas particulier du réseau en silice de période 565 nm ayant les mêmes paramètres géométriques utilisés précédemment ($b_1 = 30$ nm, $b_2 = 200$ nm, $h = 260$ nm). La figure I.20 représente le temps nécessaire pour générer numériquement une signature optique composée de 72 points (I_s et I_c calculée pour les longueurs d'onde comprises entre 350 nm et 700 nm par pas de 10 nm pour $\theta_i = 60^\circ$ et $\varphi = 0^\circ$) respectivement en fonction de M (temps T_1) lorsque N est fixé à 20 et en fonction de N (temps T_2) lorsque M est fixé à 15. Ces temps calculés correspondent aux temps CPU du calculateur utilisé (ordinateur classique Pentium IV, CPU 2GHz, RAM 1Go).

Le temps croît exponentiellement en fonction de l'ordre de troncature M et linéairement en fonction du nombre de couches N . Ainsi, il est préférable de réaliser à chaque fois l'étude de convergence permettant de déterminer les valeurs optimales de M et N réalisant un bon compromis entre précision et temps de calcul. Nous verrons plus tard l'importance de cette constatation sur les différentes méthodes de caractérisation utilisées en scatterométrie.

Dans cette partie, nous avons présenté les deux types de structures qui seront essentiellement étudiées et caractérisées dans ce manuscrit : les structures multicouches et plus particulièrement une couche déposée sur un substrat, et les réseaux de diffraction. Nous avons vu également les méthodes mathématiques permettant de calculer la signature ellipsométrique correspondante à partir des paramètres optiques et géométriques de la structure étudiée. Ce calcul sera appelé *calcul direct*.

Il ne faut pas perdre de vue que l'ellipsométrie est une méthode de caractérisation indirecte. En effet, le calcul qui permet de déterminer les paramètres recherchés d'une structure s'inscrit dans le cadre de la résolution d'un problème inverse. Dans la partie suivante, nous nous intéresserons à ce point particulier, en exposant les différentes méthodes qui peuvent être employées dans ce but.

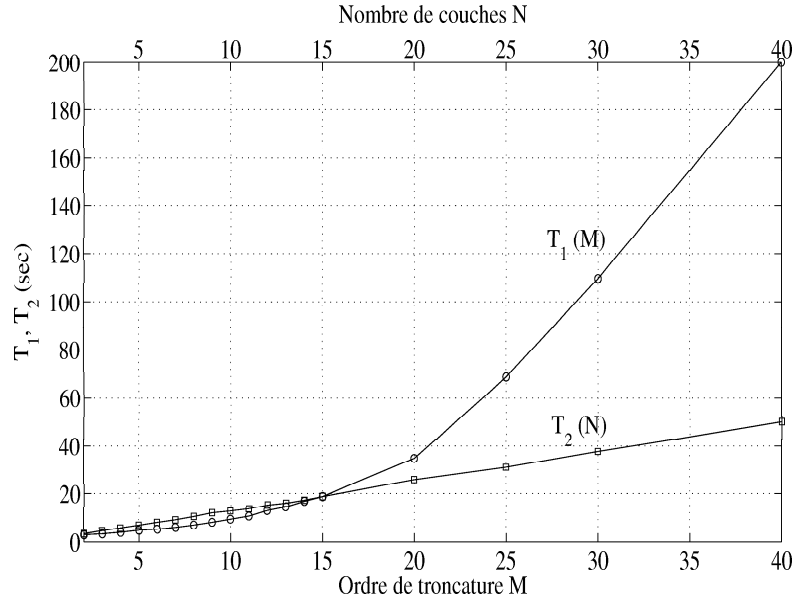


FIG. I.20: Evaluation du temps (CPU time) nécessaire pour le calcul d'une signature optique (I_s , I_c) pour ($350 \text{ nm} < \lambda < 700 \text{ nm}$) par pas de 10 nm , d'un réseau de diffraction en silice de profil trapézoïdal ($b_1 = 30 \text{ nm}$, $b_2 = 120 \text{ nm}$, $h = 180 \text{ nm}$) et de période 565 nm .

I.3 Problème inverse en ellipsométrie

La caractérisation par ellipsométrie nécessite deux étapes complètement distinctes : la mesure et le traitement. La première étape a été discutée dans le paragraphe I.1.3. Cette partie concerne le développement du second point, à savoir le traitement qui permet de déterminer les paramètres optiques et géométriques d'une structure donnée à partir de la mesure ellipsométrique effectuée sur celle-ci. Nous commencerons par donner une formulation mathématique à ce problème, puis un état de l'art assez exhaustif des différentes méthodes permettant sa résolution.

I.3.1 Position du problème

Nous avons vu précédemment que le *calcul direct* nous permettait de calculer les coefficients ellipsométriques (ψ et Δ ou I_s et I_c) pour une structure donnée (multicouche ou réseau) connaissant ses paramètres optiques et géométriques, dans des conditions expérimentales connues (angle d'incidence θ_0 , longueur d'onde λ_0 , polarisation, etc ...). En d'autres termes, ce type de calcul permet la simulation de quantités mesurables en pratique.

Le but d'une caractérisation ellipsométrique est de déterminer les paramètres optiques et géométriques de la structure étudiée qui induisent les valeurs d'intensités I_s et I_c mesurées. Il faut donc résoudre un problème inverse.

Remarque : On peut toutefois citer quelques cas simples où ces paramètres peuvent être déterminés directement, par un simple calcul analytique point par point, à partir des mesures ellipsométriques :

- le cas d'un échantillon massif non rugueux (figure I.2), où l'on cherche à déterminer l'indice de réfraction du matériau. La relation liant l'indice optique complexe (N_1)

au coefficient ellipsométrique ρ (I.14) peut s'écrire (en substituant les équations I.2, I.6 et I.7 dans I.14) :

$$\left(\frac{N_1}{N_0}\right)^2 = \sin^2 \theta_0 \left[1 + \left(\frac{1-\rho}{1+\rho}\right)^2 \tan^2 \theta_0 \right] \quad (\text{I.68})$$

Ainsi, cette équation montre que l'indice N_1 du matériau peut être déterminé si l'indice du milieu extérieur N_0 (en général égal à 1) est connu et si le rapport ellipsométrique ρ est mesuré à un angle d'incidence θ_0 donné.

- le cas d'une couche déposée sur un substrat, où l'on cherche à déterminer soit l'indice (N) et l'épaisseur (h) de la couche lorsque celle-ci est transparente ($k = 0$), soit l'indice complexe (n et k) connaissant h dans le cas d'une couche absorbante. L'indice du substrat est supposé connu. Dans ces deux cas, le système à résoudre est parfaitement déterminé : on cherche 2 paramètres (n et h ou n et k) à partir de deux mesures ψ et Δ (ou I_s et I_c) effectuées à λ et θ_0 fixés.

Afin de définir un formalisme mathématique au problème inverse appliqué à l'ellipsométrie, considérons le vecteur \vec{x} ($\vec{x} \in \mathbb{R}^p$) qui, dans le cas concret de l'ellipsométrie, représente une configuration expérimentale donnée (par exemple longueur d'onde λ , angle d'incidence θ_0). Soit y une seule observation en un point \vec{x} , qui représente une mesure ellipsométrique (I_s ou I_c) effectuée ou simulée dans la configuration expérimentale définie par \vec{x} . Le problème direct peut être formulé par l'équation suivante :

$$\begin{aligned} \mathbb{R}^p &\longrightarrow \mathbb{R} \\ \vec{x} &\longrightarrow y = F_{\vec{a}}(\vec{x}) = F(\vec{x}, \vec{a}) \end{aligned} \quad (\text{I.69})$$

où \vec{a} représente le vecteur définissant les M paramètres optiques et géométriques de la structure (par exemple n , k , h dans le cas d'une simple couche sur un substrat ou le profil d'un réseau), $F(\vec{x})$ est la fonction désignant le calcul direct parfaitement maîtrisé pour une structure \vec{a} donnée décrit dans les paragraphes I.2.1.2 et I.2.2.3.

Le but de la caractérisation par ellipsométrie est ainsi d'inférer les paramètres \vec{a} de la structure étudiée à partir des observations disponibles. La formulation du problème inverse est ainsi donnée par :

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{R}^M \\ y &\longrightarrow \vec{a} = F^{-1}(y, \vec{x}) \end{aligned} \quad (\text{I.70})$$

En général, la fonction F^{-1} est inconnue et ne possède pas de forme analytique. Ce problème est mal posé au sens de Hadamard. En effet, la résolution du problème inverse peut conduire à des solutions multiples ; c'est-à-dire que nous pourrions avoir plusieurs structures ayant différents paramètres \vec{a} mais qui donnent la même mesure y en \vec{x} . Pour résoudre le problème de la non-unicité de la solution, il est généralement souhaitable de considérer un grand nombre d'informations provenant de mesures effectuées à différentes conditions expérimentales \vec{x} . On parle dans ce cas de *régularisation mathématique*, où l'on ajoute des conditions supplémentaires que doit vérifier la solution. Par analogie avec l'équation I.70, le problème inverse sera ainsi formulé par :

$$\begin{aligned}\mathbb{R}^N &\longrightarrow \mathbb{R}^p \\ \vec{y} &\longrightarrow \vec{a} = F^{-1}(\vec{y}, \vec{x})\end{aligned}\quad (\text{I.71})$$

Il faut noter que la différence avec l'équation I.70 réside dans le fait que \vec{y} représente l'ensemble de points de mesures considérés. Par exemple, dans le cas d'une simple couche transparente on a $N = 2$ ($\vec{y} = \{I_s, I_c\}$) et $p = 2$ ($\vec{x} = \{\lambda, \theta_0\}$) ou alors $N = 4$ ($\vec{y} = \{I_s(\lambda_1), I_s(\lambda_2), I_c(\lambda_1), I_c(\lambda_2)\}$) et $p = 1$ ($\vec{x} = \theta_0$), etc ...

I.3.2 Formulation mathématique

La résolution du problème inverse est une étape purement mathématique, qui se définit comme la recherche des causes d'un phénomène à partir d'observations expérimentales. Dans notre cas, on cherche à estimer les valeurs des paramètres optiques et/ou géométriques les plus probables définissant une structure donnée dont la signature ellipsométrique calculée coïncide le plus avec celle mesurée directement sur l'échantillon.

Considérons la signature ellipsométrique (mesure de I_s et I_c pour plusieurs valeurs de λ) $\vec{y} \in \mathbb{R}^N$ où N est le nombre de mesures. Chaque observation ou mesure peut être représentée par la réalisation d'une variable aléatoire Y_i (avec $i = 1, \dots, N$). Les observations sont supposées indépendantes les unes des autres et chacune a une densité de probabilité notée f_{Y_i} . Ainsi, la probabilité que la réalisation de la variable aléatoire Y_i^m , relative à un échantillon m , prenne une valeur comprise entre y_i^m et $y_i^m + \Delta y_i^m$ est donnée par :

$$P(y_i^m < Y_i^m < y_i^m + \Delta y_i^m) = f_{Y_i}(y_i^m) \Delta y_i^m \quad (\text{I.72})$$

où Δy_i^m représente la tolérance souhaitée en chaque point de mesure entre le modèle et les mesures. Il faut noter que le modèle vrai n'existe pas. On ne se demande pas non plus si le modèle proposé est le vrai modèle, mais on cherche la probabilité de mesurer un modèle avec une certaine tolérance $\pm \Delta y_i^m$ sur chacun des points expérimentaux issus du vrai modèle.

La probabilité d'obtenir l'ensemble des observations est donnée par le produit des probabilités de l'obtenir en chaque point de mesure :

$$\begin{aligned} &P(y_1^m < Y_1^m < y_1^m + \Delta y_1^m) \cap \dots \cap P(y_N^m < Y_N^m < y_N^m + \Delta y_N^m) \\ &= P(y_1^m < Y_1^m < y_1^m + \Delta y_1^m) \times \dots \times P(y_N^m < Y_N^m < y_N^m + \Delta y_N^m)\end{aligned}\quad (\text{I.73})$$

Cette égalité résulte de l'hypothèse d'indépendance des variables aléatoires Y_i . L'idée est alors que le choix des paramètres \vec{a} est celui pour lequel cette probabilité est maximale. La fonction de vraisemblance, notée $L(\vec{y}, \vec{a})$ (de l'anglais *Likelihood*), peut être écrite par la relation suivante :

$$L(\vec{y}, \vec{a}) = \prod_{i=1}^N f_{Y_i}(y_i^m) \quad (\text{I.74})$$

Considérons une hypothèse simple, probablement la plus utilisée en pratique, où la densité de probabilité f_{Y_i} correspondante à toutes les observations est gaussienne (en supposant les erreurs systématiques négligeables). Chaque observation est affectée d'un

bruit de mesure de moyenne nulle et d'écart type σ_i . Dans ce cas, on peut écrire :

$$P(y_i^m < Y_i^m < y_i^m + \Delta y_i^m) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{y_i^m - y_i(\vec{x})}{\sigma_i} \right)^2 \right] \Delta y_i^m \quad (\text{I.75})$$

avec :

$$y_i(\vec{x}) = F(\vec{x}, \vec{a}) = E[Y_i(\vec{x})] = E[Y_i] \quad (\text{I.76})$$

L'équation I.74 devient :

$$L(\vec{y}, \vec{a}) = \prod_{i=1}^N f_{Y_i^m}(y_i^m) = \prod_{i=1}^N \left\{ \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{y_i^m - y_i(\vec{x})}{\sigma_i} \right)^2 \right] \Delta y_i^m \right\} \quad (\text{I.77})$$

Généralement, les valeurs de σ_i pour tous les points de mesure sont supposées constantes et égales, on peut ainsi écrire :

$$L(\vec{y}, \vec{a}) = \left(\frac{1}{\sigma_i \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2} \sum_{i=1}^N \left(\frac{y_i^m - y_i(\vec{x})}{\sigma_i} \right)^2 \right] (\Delta y_i^m)^N \quad (\text{I.78})$$

Ainsi, maximiser la fonction de vraisemblance L revient à minimiser l'opposé de son logarithme :

$$-\log L = -\log \left(\frac{1}{\sigma_i \sqrt{2\pi}} \right)^N - \left[-\frac{1}{2} \sum_{i=1}^N \left(\frac{y_i^m - y_i(\vec{x})}{\sigma_i} \right)^2 \right] - N \log \Delta y_i^m \quad (\text{I.79})$$

Ceci équivaut à minimiser le facteur χ^2 défini par l'équation suivante :

$$\chi^2(\vec{a}) = \sum_{i=1}^N (y_i^m - y_i(\vec{x}, a_1, \dots, a_M))^2 \quad (\text{I.80})$$

On parle dans ce cas d'ajustement par **méthode des moindres carrés**.

Il est souvent difficile d'estimer exactement l'erreur de mesure. Nous verrons plus loin une méthode simple permettant d'estimer cette erreur pour notre appareil de mesure. Dans ce cas, on cherche à minimiser le facteur χ_n^2 , appelé *facteur de mérite* ou *fonction objectif*, défini par l'équation suivante :

$$\chi_n^2(\vec{a}) = \sum_{i=1}^N \left(\frac{y_i^m - y_i(\vec{x}, a_1, \dots, a_M)}{\sigma_i} \right)^2 \quad (\text{I.81})$$

Les quantités $(y_i^m - y_i(\vec{x}, a_1, \dots, a_M))$ sont appelées les *résidus* au modèle et notées f_i , et désignent l'écart entre les points de mesure et le modèle. Lorsque ces dernières sont divisées par σ_i (équation I.81), on parle de *résidus normalisés* (d'où la notation χ_n^2 ou f_i^n).

Ainsi, le problème inverse se ramène à un problème d'optimisation paramétrique nécessitant des méthodes mathématiques efficaces et robustes. En effet, on recherche les

valeurs de \vec{a} qui minimise χ^2 (ou χ_n^2). On distingue ainsi plusieurs types d'optimisation :

- unidimensionnelle (avec un seul paramètre a_1) ou multidimensionnelle (M paramètres : a_1, \dots, a_M),
- statique (χ^2 est indépendant du temps) ou dynamique (χ^2 dépendant du temps),
- avec contraintes ou sans contraintes (égalités ou inégalités imposées sur les paramètres \vec{a}),
- stochastique (la recherche de la solution est régie par des règles probabilistes) ou déterministe (la recherche nécessite la donnée d'un point de départ \vec{a}_0 dont le déplacement itératif vers la solution est basée sur des lois déterministes).

Nous nous limiterons par la suite à l'optimisation statique et sans contrainte qui nous intéresse ici plus particulièrement.

La principale difficulté d'un problème d'optimisation est liée à la nature de la fonction à optimiser dite *fonction objectif* (dans notre cas χ^2). Pour simplifier, nous allons considérer le problème à une seule dimension, c'est-à-dire $\vec{a} \equiv a$. Dans l'espace des solutions possibles, noté $E_{sol} \in \mathfrak{R}$, un point a^* est défini comme étant un *minimum local* si :

$$\forall a \in E_{sol} / d(a, a^*) \leq \epsilon : \chi^2(a^*) \leq \chi^2(a) \quad (\text{I.82})$$

$d(a, a^*)$ désigne la distance entre le point a et a^* .

Cela signifie que dans le voisinage de a^* , il n'existe aucun point pour lequel $\chi^2(a)$ est strictement inférieur à $\chi^2(a^*)$. Dans le cas où $\chi^2(a^*) = \chi^2(a)$ pour plusieurs points a voisins de a^* , le minimum est qualifié *local faible*. Dans le cas contraire, c'est un *minimum local fort*.

Un point a^* est considérée comme un *minimum global* si :

$$\forall a, a^* \in E_{sol} : \chi^2(a^*) < \chi^2(a) \quad (\text{I.83})$$

La solution du problème correspond bien entendu au minimum global. Ces différents types de minima sont illustrés clairement sur la figure I.21. En pratique, il est primordial de choisir une méthode mathématique qui garantisse la convergence vers cette solution.

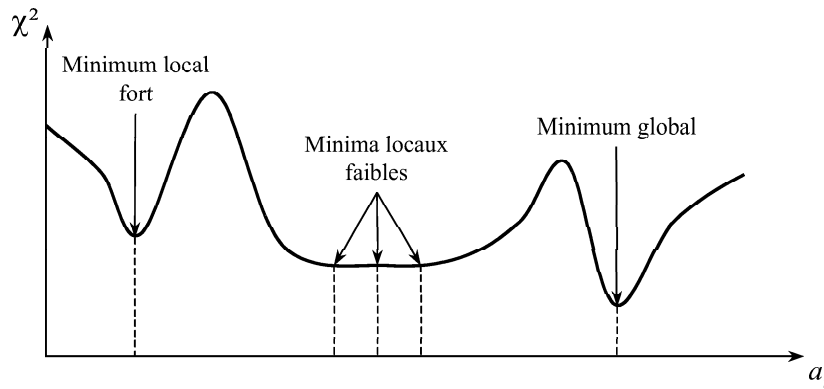


FIG. I.21: Allure schématique globale d'une fonction objectif montrant l'existence de minima locaux (fort et faible) et d'un minimum global.

Nous allons à présent illustrer ce phénomène sur un exemple réel. Pour cela, considérons le cas d'une couche de silice (SiO_2) d'épaisseur $d = 500$ nm sur un substrat de silicium (Si) dont les indices de réfraction sont connus (annexe A). Le problème consiste

à retrouver l'épaisseur d connaissant une signature ellipsométrique mesurée ($a \equiv d$) par minimisation de la quantité $\chi^2(d)$ pour un θ_0 donné. Par soucis de clarté, pour chaque angle θ_0 ($45^\circ < \theta_0 < 80^\circ$), l'erreur χ^2 est calculée en fonction de l'épaisseur d par :

$$\chi^2(d) = \sum_{i=1}^N (\vec{y}^{500} - \vec{y}(\theta_0, d))^2 \quad (\text{I.84})$$

avec \vec{y}_{500} représente la signature ellipsométrique composée des intensités $\{I_s, I_c\}$ calculée pour $400 \text{ nm} < \lambda < 800 \text{ nm}$ par pas de 10 nm et pour un angle d'incidence θ_0 . Cette signature, que l'on note \vec{y}^{500} , correspond au vecteur \vec{y}^m , et c'est λ qui est choisie comme argument pour multiplier le nombre de mesures (paragraphe I.3.1).

Nous avons étudié l'allure théorique de χ^2 en fonction du paramètre d pour différentes valeurs de θ_0 , représentée sur la figure I.22(a).

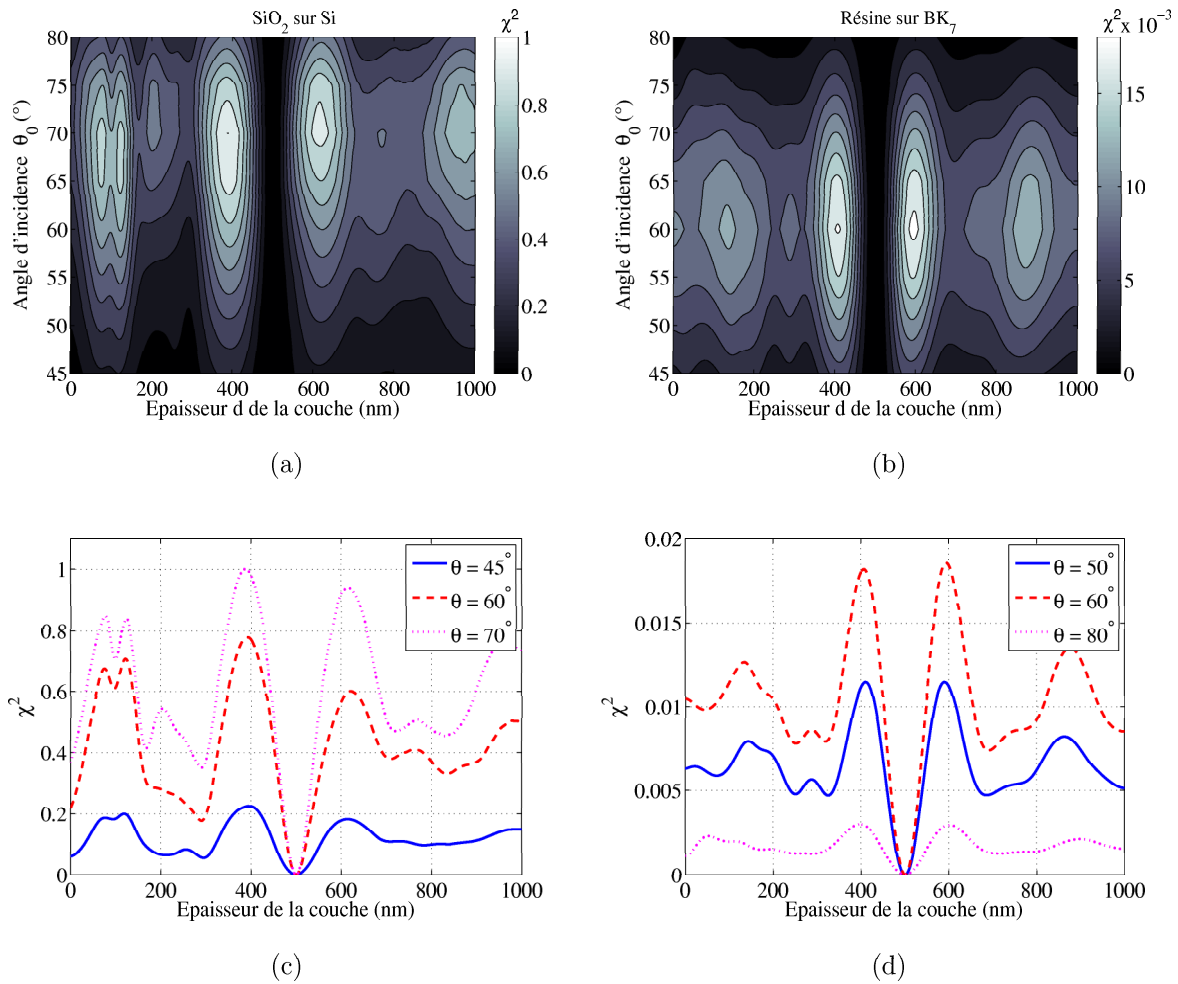


FIG. I.22: Valeur du χ^2 en fonction de l'épaisseur d pour différentes signatures optiques à λ variable paramétrées par θ_0 dans le cas (a) d'une couche de SiO_2 d'épaisseur $d = 500 \text{ nm}$ déposée sur un substrat de Si (b) d'une couche de résine de la même épaisseur déposée sur un substrat de BK_7 . (c) et (d) représentent l'erreur χ^2 extraite des figures (a) et (b) pour quelques valeurs significatives de l'angle d'incidence.

Il est évident qu'on retrouve le minimum global pour l'épaisseur $d = 500 \text{ nm}$ quelque soit l'angle d'incidence utilisé pour la signature optique. Ce minimum est nul puisque les

données utilisées sont théoriques. Toutefois, pour un angle d'incidence particulier, dans ce cas $\theta = 70^\circ$, les minima sont plus marqués que pour les autres angles (figure I.22(c)). Ces minima sont de plus en plus faibles lorsqu'on s'écarte de $\theta = 70^\circ$ proche de l'angle de Brewster pour lesquels les variations de ψ et Δ sont les plus marquées (figure I.3).

Le même phénomène est observé aussi dans le cas d'une couche de résine sur un substrat de BK_7 dans les mêmes conditions que le cas précédent. La seule différence remarquable est que le minimum global est plus marqué pour l'angle d'incidence $\theta = 60^\circ$ correspondant approximativement au nouvel angle de Brewster pour les matériaux utilisés.

La même étude est réalisée de nouveau dans le cas d'une signature optique à longueur d'onde λ fixe et angle d'incidence θ_0 variable compris entre 45° et 80° par pas de 5° .

Les valeurs du χ^2 correspondantes sont représentées sur les figures I.23(a) et I.23(b). Le minimum global théorique est obtenu pour $d = 500$ nm quelque soit la longueur d'onde utilisée. Les figures mettent en évidence l'apparition de minima locaux forts à très faible valeur de χ^2 , voisines du minimum global. L'influence de λ est très peu marquée bien qu'on décèle une amélioration pour les faibles longueurs d'onde.

L'allure périodique en épaisseur de ces courbes peut être expliquée par le fait que les intensités I_s et I_c le sont aussi. En effet, la quantité β définie par l'équation I.45 peut être écrite de façon simple dans le cas de couches transparentes sous la forme suivante [18] :

$$\beta = \frac{2\pi d}{\lambda} \sqrt{N_1^2 - N_0^2 \sin^2 \theta_0} = \frac{\pi d}{D_{\theta_0}} \quad (\text{I.85})$$

tel que

$$D_{\theta_0} = \frac{\lambda}{2\sqrt{n_1^2 - n_0^2 \sin^2 \theta_0}} \quad (\text{I.86})$$

D_θ est fonction de l'angle d'incidence θ_0 . A une longueur d'onde fixe, on peut démontrer que la fonction ρ (équation I.14) est périodique en d de période D_{θ_0} , en remplaçant r_p et r_s par leurs expressions de l'équation I.53. Ceci implique la périodicité de ψ et Δ ainsi que I_s et I_c données par les équations I.36.

Dans l'étude présentée, la signature optique des échantillons simulés est théorique et non bruitée. Or, en pratique la signature est mesurée avec une certaine incertitude. Ainsi, reprenons par exemple le cas de la figure I.23(d) pour $\lambda = 400$ nm lorsque la signature optique (I_s et I_c) est affectée d'un bruit gaussien de moyenne nulle et d'écart type $\sigma = 0,03$.

Les courbes de la figure I.24 correspondent respectivement aux erreurs χ^2 en absence et en présence du bruit dans le cas de l'échantillon en résine. Nous constatons d'une part que la valeur de χ^2 pour $d = 500$ nm n'est plus nulle. De plus, il existe d'autres points pour lesquels χ^2 possède une valeur relativement faible ($d = 224$ nm, $\chi^2 = 1,2 \cdot 10^{-3}$), ($d = 364$ nm, $\chi^2 = 1,1 \cdot 10^{-3}$), ($d = 500$ nm, $\chi^2 = 1,2 \cdot 10^{-3}$). Ceci rend difficile la détermination de la solution exacte. La prise en compte d'un bruit de mesure a tendance à niveler les courbes de χ^2 noyant le minimum global recherché au milieu d'autres minima locaux de même ordre de grandeur. Cette étude anodine nous montre l'intérêt du choix de la signature optique dans ce type de caractérisation.

En conclusion, nous pouvons constater qu' :

– à longueur d'onde λ fixe, nous avons une multiplication remarquable de minima

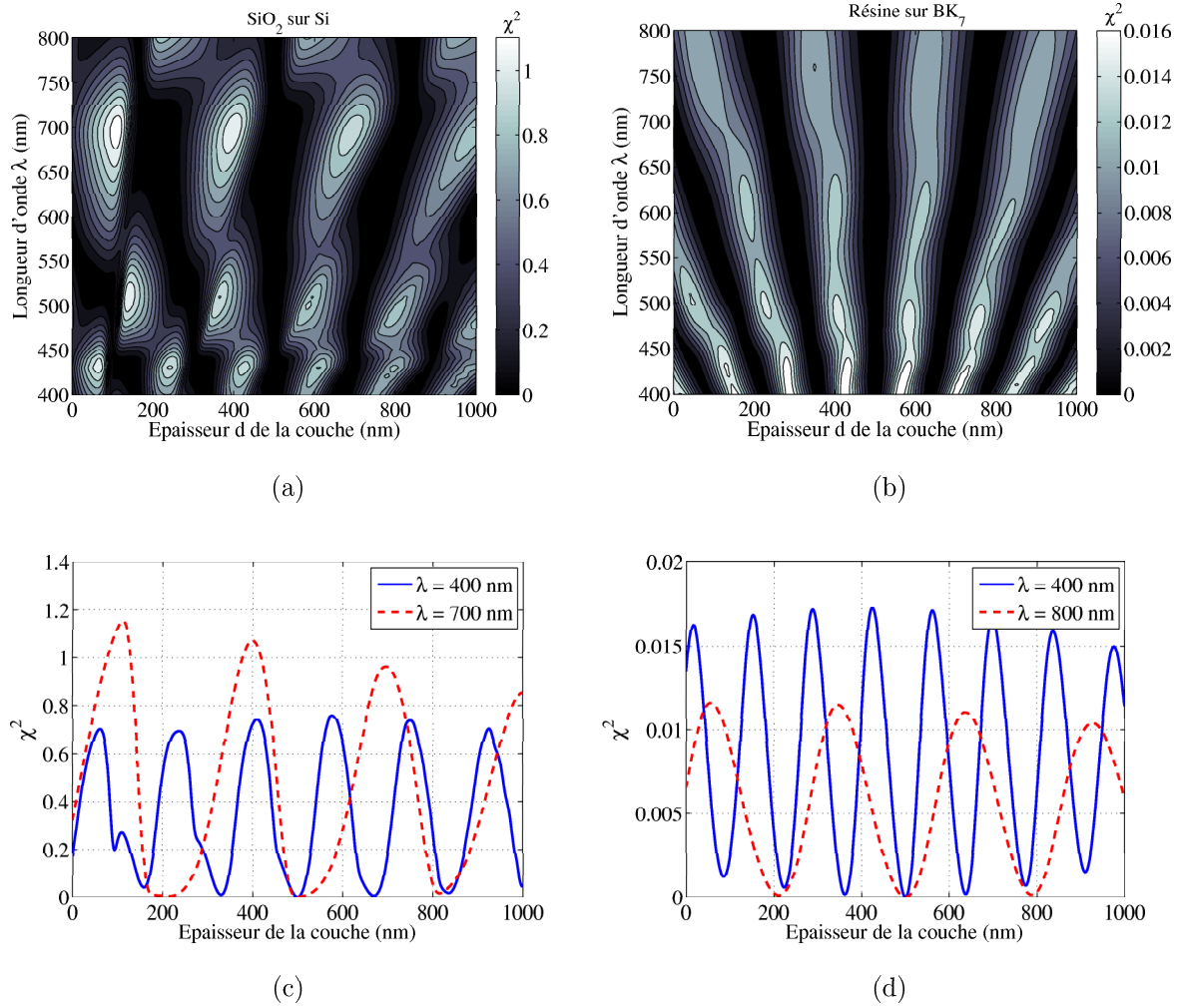


FIG. I.23: Valeur du χ^2 en fonction de l'épaisseur d pour différentes signatures optiques à θ_0 variable paramétrées par λ dans le cas (a) d'une couche de SiO₂ d'épaisseur $d = 500$ nm déposée sur un substrat de Si (b) d'une couche de résine de la même épaisseur déposée sur un substrat de BK₇. (c) et (d) représentent l'erreur χ^2 extraite des figures (a) et (b) pour quelques valeurs de la longueur d'onde.

locaux forts dont l'amplitude est du même ordre de grandeur que le minimum global ; Ce problème est accentué par la prise en compte de l'incertitude de mesure.

- à angle d'incidence θ_0 fixe, les minima locaux sont toujours présents mais moins marqués. Le minimum global est davantage marqué et ce, d'autant plus que la mesure est réalisée au voisinage de l'angle de Brewster.

Dans la suite, la signature optique à angle d'incidence θ_0 fixé proche de l'angle de Brewster sera préférentiellement utilisée.

L'optimisation constitue une véritable branche des mathématiques en perpétuelle évolution. La partie suivante donne les définitions mathématiques essentielles et parcourt quelques méthodes classiquement utilisées pour résoudre un problème d'optimisation.

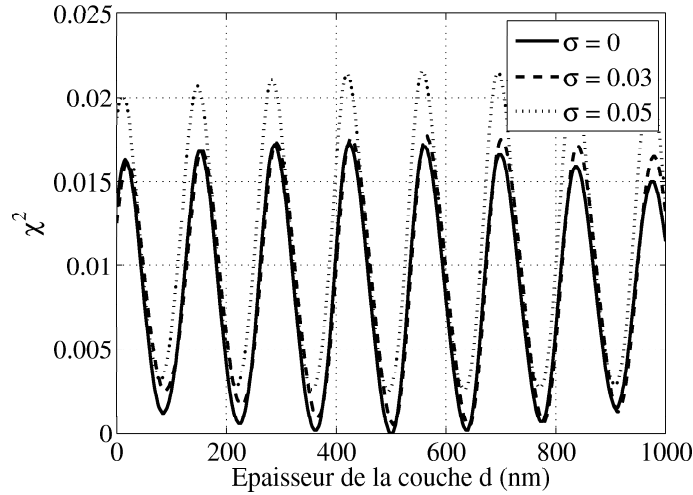


FIG. I.24: Valeurs de χ^2 , dans le cas d'une couche de résine d'épaisseur $d = 500$ nm déposée sur un substrat de BK_7 , en fonction de l'épaisseur de la couche lorsque la signature est bruitée et composée des intensités I_s et I_c pour $\lambda = 400$ nm et $45^\circ < \theta_0 < 80^\circ$ par pas de 5° pour différentes valeurs de σ .

I.3.3 Revue des méthodes d'optimisation

De façon générale, tout problème d'optimisation peut être formulé mathématiquement de la façon suivante :

$$\begin{aligned} \min f_{obj}(\vec{a}) \\ G(\vec{a}) &= 0; \\ H(\vec{a}) &> 0; \end{aligned} \quad (I.87)$$

tel que G et H sont des fonctions qui représentent les contraintes éventuelles sur \vec{a} ($\vec{a} \in \mathbb{R}^M$) et f_{obj} désigne la fonction objectif qui se trouve être dans notre cas le facteur χ^2 défini dans la partie précédente.

Dans ce qui suit, seule la minimisation de la fonction objectif f_{obj} sera considérée, puisqu'un problème de maximisation de f_{obj} peut toujours se ramener à un problème de minimisation de $-f_{obj}$.

Les méthodes d'optimisation peuvent être regroupées en trois catégories principales : les méthodes *énumératives*, les méthodes *déterministes* et les méthodes *stochastiques*.

Une seconde classification prend en compte le fait qu'elles nécessitent ou non le calcul des dérivées de la f_{obj} par rapport aux paramètres \vec{a} . Ainsi, les méthodes qui utilisent uniquement la connaissance de la fonction évaluée en chaque point \vec{a} (soit $f_{obj}(\vec{a})$), sont dites d'*ordre 0*. Elles sont appelées d'*ordre 1* si elles requièrent le calcul des dérivées premières, et d'*ordre 2* pour les dérivées secondes.

I.3.3.1 Méthodes énumératives

Le principe des méthodes énumératives est simple. La valeur de la fonction à optimiser f_{obj} est évaluée en chaque point \vec{a} de l'espace discrétisé des solutions \mathbb{R}^M . Le point qui correspond au minimum de la fonction, noté \vec{a}_{min} , est retenu comme solution.

Ces méthodes, d'ordre 0, sont facilement mises en œuvre et permettent la détermina-

tion du minimum global. Le fait d'explorer les solutions une par une, rend ces méthodes inutilisables lorsque l'espace de recherche est vaste, ce qui est souvent le cas en pratique. Elles sont essentiellement utilisées si les variables sont initialement discrètes dans un espace de recherche extrêmement limité.

I.3.3.2 Méthodes déterministes

Les méthodes déterministes ont été les plus étudiées. Elles imposent des hypothèses sur la fonction f_{obj} : celle-ci doit être continue et différentiable en tout point de l'espace \mathbb{R}^M . Au sein de cette catégorie, les méthodes se divisent en deux grandes classes : les méthodes *directes* et les méthodes *indirectes* ou *analytiques*.

Les méthodes indirectes cherchent à atteindre la solution du problème par la résolution d'un système d'équations, souvent non linéaires. Ces équations sont obtenues par l'annulation du gradient de la fonction f_{obj} :

$$\vec{\nabla} f_{obj}(\vec{a}) = \vec{0} \quad (\text{I.88})$$

La condition d'obtention d'un minimum est donnée par l'équation I.88 et l'inégalité suivante :

$$D_i > 0 \quad \text{pour } i = 1, 2, \dots, M \quad (\text{I.89})$$

tandis qu'un maximum est obtenu en vérifiant l'équation I.88 et les inégalités suivantes :

$$\begin{aligned} D_i &> 0 \quad \text{pour } i = 2, 4, 6, \dots \\ D_i &< 0 \quad \text{pour } i = 1, 3, 5, \dots \end{aligned} \quad (\text{I.90})$$

avec D_i le déterminant de la matrice hessienne défini par :

$$D_i = \begin{vmatrix} \frac{\partial^2 f}{\partial a_1 \partial a_1} & \frac{\partial^2 f}{\partial a_1 \partial a_2} & \dots & \frac{\partial^2 f_{obj}}{\partial a_1 \partial a_i} \\ \frac{\partial^2 f_{obj}}{\partial a_2 \partial a_1} & \ddots & & \frac{\partial^2 f_{obj}}{\partial a_2 \partial a_i} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f_{obj}}{\partial a_i \partial a_1} & \dots & \dots & \frac{\partial^2 f_{obj}}{\partial a_i \partial a_i} \end{vmatrix} \quad (\text{I.91})$$

Ces méthodes nécessitent la connaissance de la dérivée première et seconde de la fonction à optimiser f_{obj} . Elles ne donnent pas le type du minimum trouvé. Cependant, pour trouver le minimum global il faut calculer tous les minima solutions des équations I.88 et I.89 afin d'en extraire celui qui est le plus faible.

Les méthodes directes sont des méthodes dont le principe consiste à démarrer d'une solution potentielle et la faire évoluer dans une direction déterminée. Ainsi, la solution calculée à l'itération $(k + 1)$ peut être mise sous la forme :

$$\vec{a}^{k+1} = \vec{a}^k + \alpha^k \vec{d}^k \quad (\text{I.92})$$

où \vec{d}^k ($\in \mathbb{R}^M$) représente la direction, et α^k est un scalaire choisi tel que :

$$f_{obj}(\vec{a}^k + \alpha^k \vec{d}^k) < f_{obj}(\vec{a}^k) \quad (\text{I.93})$$

Ces méthodes diffèrent par le choix de \vec{d}^k et par la détermination du paramètre α^k ,

appelé aussi *pas de la méthode*. Les paragraphes suivants présentent de façon non exhaustive le fonctionnement de quelques unes des méthodes classiquement utilisées dans la littérature.

Méthode des plus fortes pentes [73]. C'est une méthode d'ordre 1 et l'une des plus simples. La direction de descente est choisie tel que :

$$\vec{d}^k = -\vec{\nabla} f_{obj}(\vec{a}^k) \quad (\text{I.94})$$

Le principal avantage de cette méthode est sa facilité de mise en œuvre. Par contre, elle n'est pas très efficace en pratique dû à sa faible vitesse de convergence. En effet, elle converge lentement au voisinage de l'optimum et rapidement loin de celui-ci. De plus, elle ne prend pas en considération l'allure de la fonction coût.

Méthode du gradient conjugué. Cette méthode d'ordre 1 est très proche de la méthode des plus fortes pentes, avec une certaine amélioration. Elle repose sur le concept de directions conjuguées alors que les gradients successifs sont orthogonaux entre eux et aux directions précédentes. Ceci a permis de réduire considérablement le temps de convergence.

La méthode de Fletcher et Reeves [74] et la méthode de Polak et Ribière [75] sont deux variantes de la méthode du gradient conjugué. Dans la première, la direction de descente est donnée par :

$$\vec{d}^{k+1} = -\vec{\nabla} f_{obj}(\vec{a}^k) + \beta^k \vec{d}^k \quad (\text{I.95})$$

avec

$$\beta^k = \frac{\left[\vec{\nabla} f_{obj}(\vec{a}^{k+1}) \right]^T \vec{\nabla} f_{obj}(\vec{a}^{k+1})}{\left[\vec{\nabla} f_{obj}(\vec{a}^k) \right]^T \vec{\nabla} f_{obj}(\vec{a}^k)} \quad (\text{I.96})$$

La deuxième s'est développée quelques années plus tard. Elle reprend le principe de la méthode de Fletcher et Reeves en remplaçant le paramètre β_k par :

$$\beta^k = \frac{\left[\vec{\nabla} f_{obj}(\vec{a}^{k+1}) - \vec{\nabla} f_{obj}(\vec{a}^k) \right]^T \vec{\nabla} f_{obj}(\vec{a}^{k+1})}{\left[\vec{\nabla} f_{obj}(\vec{a}^k) \right]^T \vec{\nabla} f_{obj}(\vec{a}^k)} \quad (\text{I.97})$$

Cette variante est réputée plus performante et sa convergence est beaucoup plus rapide que celle de Fletcher et Reeves.

Méthode de Gauss-Newton. C'est une méthode qui est souvent employée pour résoudre les problèmes d'optimisation. Elle fait appel à la matrice Hessienne H de la f_{obj} et, de ce fait, est qualifiée de méthode du second ordre. Au voisinage immédiat du point courant \vec{a}^k , la fonction est supposée quadratique et donnée par la formule de Taylor :

$$f_{obj}(\vec{a}^{k+1}) = f_{obj}(\vec{a}^k) + \vec{\nabla}(f_{obj}(\vec{a}^k))^T \cdot \vec{s}^k + \frac{1}{2}(\vec{s}^k)^T \cdot H \cdot \vec{s}^k \quad (\text{I.98})$$

avec

$$\vec{s}^k = \vec{a}^{k+1} - \vec{a}^k = \alpha^k \vec{d}^k \quad (\text{I.99})$$

Le minimum est défini par la condition suivante :

$$\vec{\nabla} f_{obj}(\vec{a}^{k+1}) = 0 \quad (\text{I.100})$$

ce qui donne, en négligeant les termes croisés d'ordre 2 et ceux d'ordre 3 :

$$\vec{s}^k = - [\nabla^2 f_{obj}(\vec{a}^k)]^{-1} \cdot \vec{\nabla} f_{obj}(\vec{a}^k) = -H^{-1} \cdot \vec{\nabla} f_{obj}(\vec{a}^k) \quad (\text{I.101})$$

soit :

$$\vec{d}^k = \frac{-H^{-1} \cdot \vec{\nabla} f_{obj}(\vec{a}^k)}{\alpha^k} \quad (\text{I.102})$$

Le fait d'utiliser la matrice Hessienne, souvent difficile et longue à calculer, rend cette méthode peu pratique.

Méthode de Quasi-Newton. Cette méthode, d'ordre 2, ne calcule pas la matrice H^{-1} à chaque itération ; elle la construit itérativement à partir de celle calculée précédemment.

La méthode de Quasi-Newton la plus connue est celle de BFGS [76, 77, 78, 79] (*Broyden-Fletcher-Goldfarb-Shanno*). Dans cet algorithme, l'expression de la matrice Hessienne (H_{k+1}^{-1}) à l'itération $k+1$ est donnée par la formule de récurrence suivante :

$$(H^{-1})^{k+1} = (H^{-1})^k - \frac{(H^{-1})^k \vec{s}^k [(H^{-1})^k \vec{s}^k]^T}{(\vec{s}^k)^T (H^{-1})^k \vec{s}^k} + \frac{\vec{q}^k (\vec{q}^k)^T}{(\vec{q}^k)^T \vec{s}^k} + \phi^k [(\vec{s}^k)^T (H^{-1})^T \vec{s}^k] \vec{v}^k (\vec{v}^k)^T \quad (\text{I.103})$$

avec

$$\vec{v}^k = \left[\frac{\vec{q}^k}{(\vec{q}^k)^T \vec{s}^k} - \frac{(H^{-1})^k \vec{s}^k}{(\vec{s}^k)^T (H^{-1})^k \vec{s}^k} \right]$$

$$\vec{q}^k = \vec{\nabla} f_{obj}(\vec{a}^{k+1}) - \vec{\nabla} f_{obj}(\vec{a}^k)$$

et ϕ^k est fixé à 0 ou 1.

Il faut toutefois noter que ces méthodes ne sont efficaces que si l'approximation quadratique (équation I.98) de la fonction f_{obj} autour de l'optimum recherché est vérifiée.

Méthode de Levenberg-Marquardt. L'algorithme de Levenberg-Marquardt (LM) [80, 81] est devenue une technique standard, employée dans diverses disciplines scientifiques, pour la résolution des problèmes d'ajustement au moindre carrés de modèles non linéaires.

Levenberg et Marquardt ont proposé un algorithme permettant de combiner la méthode de Gauss-Newton (équation I.101) et la méthode des plus fortes pentes (équation I.94), suivant que f_{obj} se rapproche d'une forme quadratique ou non. Loin de la solution, il se comporte comme une méthode des plus fortes pentes, alors qu'au voisinage de la solution il devient une méthode de Gauss-Newton.

Reprenons l'expression I.81 de χ^2 (ou f_{obj}). En dérivant deux fois successivement, la matrice Hessienne est obtenue par :

$$\frac{\partial^2 f_{obj}}{\partial a_k \partial a_l} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[\frac{\partial y_i(\vec{x}, \vec{a})}{\partial a_k} \frac{\partial y_i(\vec{x}, \vec{a})}{\partial a_l} - [y_i^m - y_i(\vec{x}, \vec{a})] \frac{\partial^2 y_i(\vec{x}, \vec{a})}{\partial a_l \partial a_k} \right] \quad (\text{I.104})$$

En négligeant les termes de second ordre, l'équation précédente s'écrit :

$$\frac{\partial^2 f_{obj}}{\partial a_k \partial a_l} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[\frac{\partial y_i(\vec{x}, \vec{a})}{\partial a_k} \frac{\partial y_i(\vec{x}, \vec{a})}{\partial a_l} \right] \quad (\text{I.105})$$

On peut en simplifiant poser :

$$\beta_k = -\frac{1}{2} \frac{\partial f_{obj}}{\partial a_k} \quad (\text{I.106})$$

$$u_{kl} = \frac{1}{2} \frac{\partial^2 f_{obj}}{\partial a_k \partial a_l} \quad (\text{I.107})$$

ou sous forme matricielle :

$$\vec{\beta} = -\frac{1}{2} \vec{\nabla} f_{obj} \quad [U] = \frac{1}{2} H \quad (\text{I.108})$$

où U est appelée la matrice de courbure.

L'algorithme de LM repose sur la définition d'une nouvelle matrice U' exprimée par :

$$[U'] = [U] + \lambda I_d \quad (\text{I.109})$$

avec I_d une matrice identité de dimension $M \times M$.

Ainsi, la nouvelle direction \vec{s} à l'itération k (I.99) sera donnée par l'équation :

$$[U'] \vec{s} = \vec{\beta} \quad (\text{I.110})$$

Lorsque λ est très grand $[U']$ est à diagonale dominante, l'équation I.110 devient :

$$s_l = \frac{1}{\lambda u_{ll}} \beta_l \quad (\text{I.111})$$

On retrouve le schéma I.94 des plus fortes pentes .

Dans le cas contraire, lorsque λ tend vers 0 l'équation I.110 équivaut au schéma I.101 d'une méthode de Gauss-Newton.

L'algorithme de LM est considéré comme le plus efficace en pratique. Cette méthode sera employée par la suite en comparaison avec la méthode par réseaux de neurones.

Méthode du simplex. C'est Nelder et Mead [82] les fondateurs de cette méthode. Contrairement aux autres méthodes qui démarrent généralement d'un point initial, cette dernière utilise un simplex de départ qui est une figure géométrique régulière à $M + 1$ dimensions dans un espace à M dimensions (triangle dans un espace à 2 dimensions, tétraèdre dans un espace à 3 dimensions, etc...). Pour chaque point formant le simplex, la fonction objectif est évaluée. Le sommet à la plus forte évaluation est supprimé et un nouveau sommet est défini par une transformation de la figure en appliquant des simples opérations algébriques sur les points, qui se traduisent par des transformations géométriques (réflexion, contraction, expansion, etc). Cette opération se répète jusqu'à ce qu'un des sommets converge vers le minimum de la fonction objectif.

Cette méthode est facile à mettre en œuvre, n'est pas gourmande en terme de mémoire et ne nécessite pas le calcul de dérivées de la fonction coût. Par contre, l'un de ses points faibles est le temps de calcul qui est non négligeable puisqu'elle nécessite l'évaluation de la fonction à optimiser un grand nombre de fois. De plus, elle est sensible aux minima locaux et donc fortement dépendante du point de départ.

I.3.3.3 Méthodes évolutionnistes

Les méthodes stochastiques appartiennent à la famille des méthodes où la recherche de la solution est guidée par un processus aléatoire. Lorsque l'on veut résoudre des problèmes de grandes dimensions, les méthodes traditionnelles ne s'avèrent pas efficaces et surtout peu robustes. On préfère dans ce cas avoir recours à ce genre de méthodes. Dans la suite, nous exposerons brièvement le principe de base de deux d'entre elles : la méthode du recuit simulé et les algorithmes génétiques. Toutefois, la littérature scientifique regroupe de nouvelles méthodes qualifiées de métaphore biologique (essaim d'abeille, colonie de fourmis, etc. ...) encore au stade expérimental à l'heure actuelle et trop peu utilisées pour être développées ici.

Méthode de recuit simulé. La méthode du recuit simulé est inspirée du processus utilisé en métallurgie [83]. Elle est issue d'une analogie entre le phénomène physique de refroidissement lent d'un corps en fusion, qui le conduit à un état solide stable présentant un minimum d'énergie.

Cette méthode d'optimisation exploite le critère défini par Métropolis [84] qui permet de décrire l'évolution de l'équilibre thermodynamique. La fonction objectif à minimiser étant l'énergie E du système, une solution \vec{a} peut être considérée comme un état donné de la matière dont $f_{obj}(\vec{a})$ est l'énergie E . A partir d'une solution initiale \vec{a}_0 , on génère aléatoirement une solution \vec{a}_1 quelconque afin d'obtenir un état différent de celui de départ. Cette solution est acceptée si la fonction objectif f_{obj} respecte la condition suivante :

$$\Delta f_{obj} = f_{obj}(\vec{a}_1) - f_{obj}(\vec{a}_0) < 0 \quad (\text{I.112})$$

Sinon, c'est à dire dans le cas où $\Delta f_{obj} \geq 0$, l'état voisin peut tout de même être accepté avec une probabilité p défini par :

$$p = \exp\left(\frac{-\Delta f_{obj}}{T}\right) \quad (\text{I.113})$$

L'acceptation d'une mauvaise solution permet d'explorer une région plus grande dans l'espace des solutions $E_{sol} \in \mathbb{R}^M$ et tend à éviter de s'enfermer trop vite dans la recherche d'un optimum local. Le paramètre de contrôle est la température T du système. Il influe sur la probabilité p d'accepter une mauvaise solution et son rôle est très important au cours du processus. Si T décroît fortement, l'algorithme risque d'être piégé rapidement dans un minimum local, alors que si cette décroissance est faible ceci entraîne une convergence très lente. Une solution courante consiste à utiliser une variation de température logarithmique.

Polgar et al. [85] ont proposé l'utilisation de la méthode du recuit simulé pour résoudre le problème inverse en ellipsométrie afin de caractériser une couche de silicium poreux. L'algorithme a ensuite été amélioré pour favoriser la rapidité et améliorer la convergence [86]. A notre connaissance, cette méthode n'a pas été utilisée en scatterométrie pour

déterminer le profil géométrique des réseaux de diffraction à partir d'une signature ellipsométrique.

Les algorithmes génétiques. Les algorithmes génétiques sont inspirés du fonctionnement de l'évolution naturelle élaborée par Charles Darwin. Après avoir étudié les systèmes évolutifs dans les années 1960, J. Holland exposait les principes fondamentaux des algorithmes génétiques en 1975 [87]. Ils ont ensuite été popularisés par son élève Goldberg [88].

Dans cette technique, on simule le processus d'évolution d'une population d'individus. On part d'une population initiale composée d'un certain nombre de solutions potentielles choisies aléatoirement et représentant les individus. C'est la première génération ou la population parent. La fonction $f_{obj}(\vec{a})$ représente le degré d'adaptation d'un individu à l'environnement. Chaque individu appartenant à la population initiale est ensuite évalué par la fonction f_{obj} . L'étape suivante consiste à créer la prochaine génération en utilisant les "meilleurs individus" en terme de f_{obj} de la population précédente. Cet étape fait intervenir la *sélection* stochastique d'un ou deux parents, puis des opérateurs génétiques (*croisement* et *mutation*) pour la génération de nouveaux individus. Le processus est répété sur plusieurs générations jusqu'à ce qu'on vérifie un critère d'arrêt. Au final, la population produite est riche en individus mieux adaptés, et proches de la solution optimale.

Ces algorithmes sont facilement adaptables à différents types de problèmes avec plusieurs paramètres à estimer et ne nécessitent pas le calcul de la dérivée (première ou seconde) de la fonction à optimiser. Leur avantage majeur est leur capacité à déterminer le minimum global dans un délai plus ou moins raisonnable.

L'introduction des algorithmes génétiques dans le domaine d'optique reste relativement récente. Néanmoins, ils ont montré une capacité remarquable pour la caractérisation optique et géométrique des couches minces à partir des données ellipsométriques [89]. Dans une partie de sa thèse soutenue en 2003 [90], S. Robert a montré l'efficacité des algorithmes génétiques pour la caractérisation géométriques des réseaux de diffraction, comparativement à d'autres méthodes plus classiques comme celle de Levenberg-Marquardt. La signature optique utilisée, consistait en différentes efficacités de diffraction mesurées à angle d'incidence variable, pour une longueur d'onde fixe. L'inconvénient majeur réside dans le temps de calcul assez long dû à la complexité du problème direct de diffraction (paragraphe I.2.2.3) fortement sollicité par les algorithmes génétiques.

A part les quelques méthodes stochastiques, la plupart des méthodes présentées sont confrontées à l'existence de minima locaux. Afin de garantir la convergence vers la solution du problème, l'utilisateur doit fournir un point de départ voisin ou relativement proche de la solution. Ceci reste l'inconvénient majeur du mode de fonctionnement des méthodes classiques bien que celles-ci soient très employées. L'un des objectifs principaux de cette thèse est de développer une méthode qui peut surmonter ces difficultés et constituer un outil efficace pour la caractérisation en optique.

I.3.4 Estimation de l'intervalle de confiance

Lorsque les paramètres d'un modèle sont estimés, dans notre cas par la résolution du problème inverse, il est utile et primordial de juger la pertinence du résultat. Cela

signifie qu'il est indispensable de donner la marge d'erreur associée à l'estimation des paramètres, ce que l'on appelle aussi intervalle de confiance. Dans ce paragraphe, nous allons résumer la façon de calculer un intervalle de confiance associé aux résultats donnés par une méthode d'optimisation classique décrite précédemment (section I.3.3).

L'intervalle de confiance d'un paramètre a_i , noté Δa_i , est l'intervalle dans lequel est incluse la vraie valeur a_i^{vrai} du paramètre.

$$a_i - \Delta a_i < a_i^{vrai} < a_i + \Delta a_i \quad (\text{I.114})$$

Il est souvent estimé en fonction d'un *facteur de risque* α . Ainsi, on définit un intervalle de confiance de niveau $(1 - \alpha)$ lorsqu'il contient la valeur du paramètre à estimer avec une probabilité $(1 - \alpha)$, d'où :

$$P [a_i - \Delta a_i < a_i^{vrai} < a_i + \Delta a_i] = 1 - \alpha \quad (\text{I.115})$$

Pour $\alpha = 0,05$ par exemple, la probabilité de trouver la vraie valeur dans l'intervalle de confiance est de 0,95. Rappelons que le paramètre a_i qu'on estime peut être vu comme la réalisation d'une variable aléatoire notée A_i de distribution gaussienne d'écart type $\sigma(a_i)$. Ainsi, les valeurs de $(1 - \alpha)$ exprimées en % communément utilisées sont 68,3%, 95,4% ou 99,73% correspondant respectivement à un intervalle défini 1σ , 2σ ou 3σ .

$$P [a_i - z \cdot \sigma(a_i) < a_i^{vrai} < a_i + z \cdot \sigma(a_i)] = 1 - \alpha \quad (\text{I.116})$$

où z est le fractile d'ordre $(1 - \alpha)/2$ de la loi normale, et $\sigma(a_i)$ est l'écart type de la variable aléatoire A_i de moyenne a_i .

Le paramètre $\sigma(a_i)$ étant inconnu, il est possible de le remplacer par un estimateur sans biais noté $s(a_i)$ défini par n tirages de la variable aléatoire A_i :

$$s(a_i) = \frac{1}{n-1} \sum_{i=1}^n (a_i - a_i^{vrai})^2 \quad (\text{I.117})$$

Dans ce cas, la loi normale ne s'avère plus un bon candidat. En effet, l'utilisation de $s(a_i)$ introduit une source supplémentaire d'erreur particulièrement si $(n - 1)$ est petit. Il faut ainsi remplacer la valeur z par une valeur plus adéquate pour tenir compte de l'erreur supplémentaire. Le meilleur choix est la loi de Student (figure I.25). L'équation I.116 peut ainsi s'écrire :

$$P [a_i - t \cdot s(a_i) < a_i^{vrai} < a_i + t \cdot s(a_i)] = 1 - \alpha \quad (\text{I.118})$$

où t est le fractile d'ordre $(1 - \alpha)/2$ de la loi de Student.

Remarque : Dans la pratique, lorsque σ n'est pas connu, la loi de Student est utilisée seulement lorsque ν est petit (< 100). En effet, lorsque ν devient très grand, la loi de Student tend vers la loi normale et les expressions I.116 et I.118 sont identiques.

Comme la valeur vraie a_i^{vrai} n'est généralement pas connue, $s(a_i)$ n'est pas calculable par I.117. Dans notre cas, sachant que a_i est estimé à partir des mesures y_i , $\sigma^2(a_i)$ peut

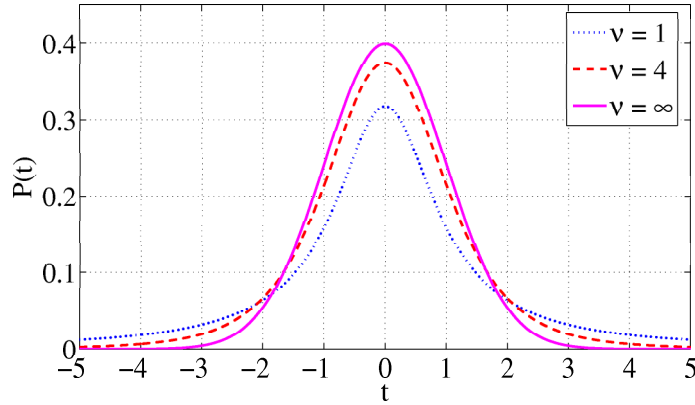


FIG. I.25: Distributions de student à ν degrés de liberté, représentation de la distribution normale ($\nu = \infty$)

être exprimée par la loi de propagation des erreurs de la façon suivante :

$$s^2(a_i) = \sum_{j=1}^N \sigma_j^2 \left(\frac{\partial a_i}{\partial y_j} \right)^2 \quad (\text{I.119})$$

On peut démontrer que si les σ_j sont tous égaux $\forall j$ [91] :

$$s^2(a_i) = \sigma_i^2 C_{ii} \quad (\text{I.120})$$

avec C est la matrice de variance-covariance des paramètres \vec{a} . définie par :

$$[C] = [U]^{-1} = \left[\frac{1}{2} J^t J \right]^{-1} = 2 [H]^{-1} \quad (\text{I.121})$$

où J est la matrice jacobienne des paramètres définie par :

$$\begin{pmatrix} \frac{\partial f_1}{\partial a_1} & \cdots & \frac{\partial f_N}{\partial a_1} \\ \frac{\partial f_1}{\partial a_2} & \cdots & \frac{\partial f_N}{\partial a_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial a_M} & \cdots & \frac{\partial f_N}{\partial a_M} \end{pmatrix} \quad (\text{I.122})$$

avec $f_i = y_i - y_i(\vec{x}, a_1, \dots, a_M)$ les résidus.

Remarque : L'expression de $\sigma^2(a_i)$ (équation I.120) dépend du choix de χ^2 . Dans notre cas, nous avons utilisé χ^2 défini dans l'équation I.80. Dans le cas où l'on emploie χ_n^2 normalisé (équation I.81), on obtient $\sigma^2(a_i) = C_{ii}$.

Le paramètre σ_j figurant dans l'équation I.120, désignant l'écart type sur les mesures y_j , est généralement inconnu. Si on suppose que les σ_i sont égaux $\forall i$, une estimation non biaisée peut être donnée par :

$$s^2(y_i) = \frac{1}{N - M} \sum_{j=1}^N (y_j - y_j(\vec{x}, a_1, \dots, a_M))^2 \quad (\text{I.123})$$

$N - M = \nu$ désigne le nombre de degrés de liberté.

En substituant l'équation I.80 dans I.123, $s^2(y^i)$ devient :

$$s^2(y_j) = \frac{\chi^2}{N - M} \quad (\text{I.124})$$

Finalement, l'intervalle de confiance s'exprime par :

$$\Delta a_i = ts(a_i) = t\sqrt{\sigma_j^2 C_{ii}} = t\sqrt{C_{ii}}\sqrt{\frac{\chi^2}{N - M}} \quad (\text{I.125})$$

Pour finir, estimons brièvement le bruit de mesure de notre ellipsomètre. A titre d'exemple, nous avons effectué 10 mesures successives des spectres d'intensités I_s et I_c dans les mêmes conditions expérimentales ($\theta = 60^\circ$, $400 \text{ nm} < \lambda < 800 \text{ nm}$) pour deux échantillons différents constitués d'une couche de résine déposée sur un substrat de verre BK_7 , tels que les épaisseurs de la résine sont respectivement de l'ordre de 350 nm (E_{R_1}) et 800 nm (E_{R_2}). Nous avons ensuite estimé les écarts-type σ_i des mesures en chaque point, reportés dans la figure I.26 pour les deux échantillons. Les valeurs moyennes des intensités I_s et I_c sont également reportées sur la figure I.27.

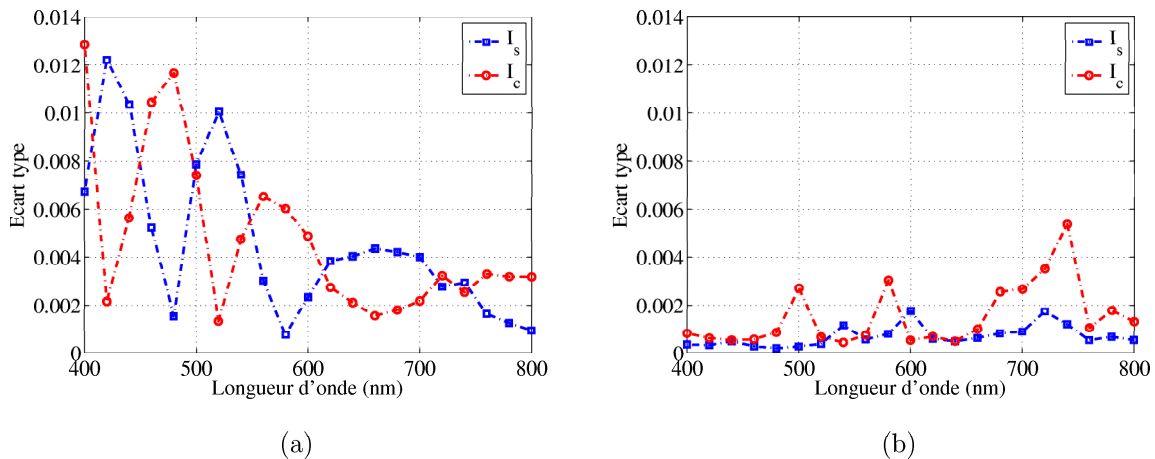


FIG. I.26: Ecart type des intensités I_s et I_c estimé à partir de 10 mesures expérimentales effectuées sur deux échantillons constitués d'une couche de résine d'épaisseur (a) 350 nm noté E_{R_1} et (b) 800 nm noté E_{R_2} , déposée sur un substrat de verre BK_7 .

Ces courbes nous indiquent que l'erreur de mesure n'est pas constante et dépend dans ce cas de la longueur d'onde pour un échantillon donné. De plus, elle varie selon la structure analysée et donc, il est difficile d'établir une loi générale permettant de déterminer sa valeur exacte.

Toutefois, il est important d'étudier l'importance de la connaissance précise de σ_i sur les résultats. Pour cela, nous avons réalisé l'étude suivante sur l'échantillon E_{R_1} . La caractérisation des propriétés optiques et géométriques de la couche de résine est d'abord réalisée par une optimisation classique (Levenberg-Marquardt). Les erreurs de mesure σ_i sont prises dépendantes de λ et égales à celles calculées sur la figure I.26(a). Dans ce cas, l'optimisation consiste à minimiser le facteur χ_n^2 défini par l'équation I.81. Le résultat de l'optimisation est affichée sur la figure I.28 (σ mesuré), montrant un bon ajustement entre

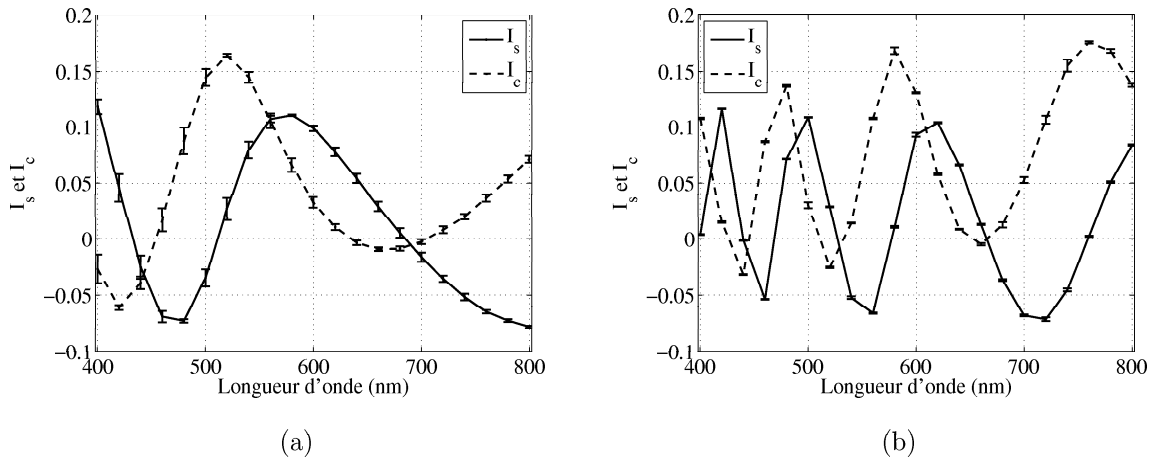


FIG. I.27: Valeurs moyennes des intensités I_s et I_c calculés à partir de 10 mesures expérimentales effectuées sur deux échantillons constitués d'une couche de résine d'épaisseur (a) 350 nm E_{R1} et (b) 800 nm E_{R2} , déposée sur un substrat de verre BK₇.

les intensités calculées et celles mesurées. Les paramètres optogéométriques correspondant calculés sont donnés dans le tableau I.2.

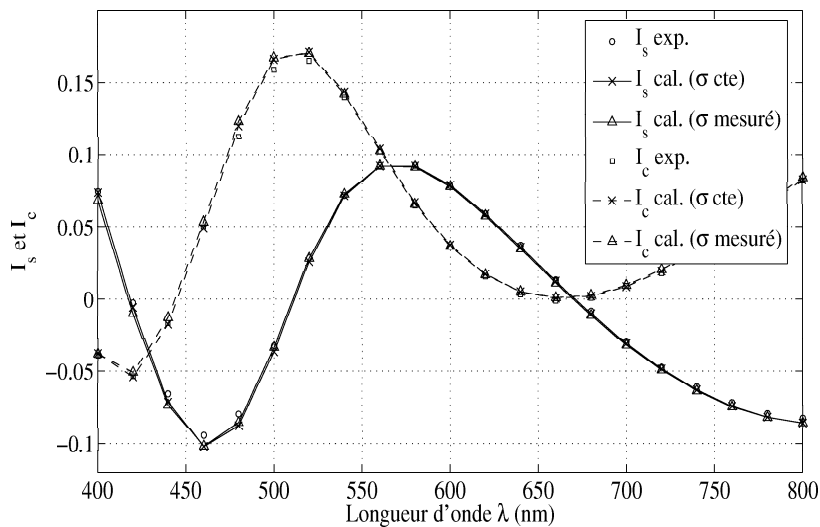


FIG. I.28: Ajustement des intensités expérimentales I_s et I_c par l'algorithme de Levenberg-Marquardt, dans les deux cas où σ est prise constante et celui où σ est mesurée expérimentalement.

	A	B (nm ²)	C (nm ²)	h (nm)
σ mesuré	1,606	0,92	0,47	364,3
σ cte	1,606	0,85	0,62	364,7

TAB. I.2: Paramètres A, B, C et h de l'échantillon E_{R1} déterminés par l'algorithme de Levenberg-Marquardt en considérant un bruit mesuré (σ mesuré) et un bruit fixe (σ cte).

En pratique, cette démarche rigoureuse est longue et fastidieuse.

Nous allons à présent considérer que les erreurs de mesure sont constantes, et refaire l'optimisation en utilisant cette fois-ci le facteur χ^2 non normalisé ($\sigma = 1$). Le résultat est affiché sur la figure I.28 (σ cte), et les paramètres optogéométriques estimés sont donnés dans le tableau I.2. Le résultat de l'optimisation ainsi que les paramètres optogéométriques sont relativement proches dans les deux cas considérés. A l'aide de la formule I.123, nous pouvons à présent estimer la valeur de σ [91] supposée constant pour toutes les mesures. La valeur trouvée est $\sigma = 0.0033$, qui constitue une estimation de l'erreur de mesure. La même procédure est réalisée pour le second échantillon E_{R_2} , et le résultat se trouve identique pour σ . Par la suite, l'erreur de mesure sera ainsi fixé à $\sigma = 0,0033$.

Dans cette partie, nous avons présenté le problème inverse en ellipsométrie qui s'apparente à une optimisation paramétrique. Le formalisme mathématique correspondant a été défini. Les différentes méthodes, classiquement utilisées pour résoudre ce type de problème ont été exposées. Nous nous sommes concentrés sur les avantages et les inconvénients de chacune d'elles. Nous avons terminé cette partie par l'estimation de l'intervalle de confiance associé aux paramètres recherchés.

I.4 Conclusion

Ce chapitre a présenté la caractérisation ellipsométrique. Les différents modèles de structures, principalement les couches minces déposées sur un substrat et les réseaux de diffraction, qui feront l'objet de la caractérisation ont été détaillés. La première étape consiste à connaître la réponse ellipsométrique de ces structures à partir des propriétés optiques et géométriques, ainsi que les conditions expérimentales choisies. Ce calcul direct est simple et facile à mettre en œuvre dans le cas des structures multicouches à travers les lois fondamentales de l'optique géométrique, ce qui n'est pas le cas pour les réseaux de diffraction. En effet, la reformulation mathématique du problème électromagnétique des réseaux est beaucoup plus complexe, nécessitant la résolution numérique des équations de Maxwell dans les différentes régions de l'espace. Dans le cadre de notre travail, nous avons opté pour la méthode modale par développement de Fourier (MMMFE). Cette méthode est la plus utilisée aujourd'hui et permet de résoudre la majorité des cas rencontrés en pratique. Certaines précautions sont également à prendre en compte au niveau du choix de certains paramètres, comme l'ordre de troncature M pour la série de Fourier et le nombre N de couches décrivant le profil géométrique du réseau. Nous avons mis en évidence le fait que ce choix dépende de la dimension (période) et de la forme des réseaux ainsi que de leurs propriétés optiques (indice de réfraction).

La méthode ellipsométrique est un outil de caractérisation indirect, nécessitant deux phases : la mesure et le traitement. Dans la dernière partie de ce chapitre, nous avons revu les différentes méthodes d'optimisation capables de traiter une mesure ellipsométrique afin d'en extraire les propriétés recherchées.

Dans le chapitre suivant, nous présenterons un outil mathématique original pour réaliser ce type de traitement : les réseaux de neurones. Nous montrerons sur des exemples simples les avantages que possèdent ces derniers par rapport aux méthodes classiques d'optimisation.

CHAPITRE II

Réseaux de neurones

Le présent chapitre porte sur un outil mathématique complexe dont le fonctionnement est inspiré du cerveau humain : le réseau de neurones (RN). Afin d'éviter de noyer le lecteur dans des notions mathématiques, nous illustrerons au fur et à mesure de la présentation des différentes étapes de fonctionnement d'un réseau de neurones, comment cet outil peut être employé en métrologie optique, à travers des exemples réels portant sur la caractérisation ellipsométrique de couches minces de résine photosensible.

Le RN est un système mathématique possédant une capacité d'apprentissage de l'environnement extérieur. Il est capable de modéliser un processus physique à partir de données expérimentales mises à disposition afin de généraliser sur de nouvelles données. Dans notre cas, le RN sera essentiellement utilisé pour apprendre la relation liant la signature ellipsométrique (entrée du RN), aux paramètres optiques et géométriques d'une structure donnée (sorties du RN). En d'autres termes, son rôle sera de résoudre le problème inverse en ellipsométrie.

Dans un premier temps, le principe de base des réseaux de neurones artificiels est présenté en exposant plus particulièrement les réseaux à couches appelés perceptrons multicouches (PMC). Nous détaillerons ensuite le processus d'apprentissage. Nous verrons comment les intervalles de confiance associés aux paramètres de sortie peuvent être estimés par l'intermédiaire d'un second RN.

La partie suivante est consacrée à la réduction de la complexité du RN en employant une méthode particulière (HVS, en anglais Heuristic Variable Selection), dont nous détaillerons le fonctionnement.

Dans la dernière partie du chapitre, nous présenterons une fonctionnalité particulière : la classification. On montre que les réseaux de neurones sont capables de trouver la même solution que celle fournie par la règle de Bayes, sans aucune condition particulière. En effet, la sortie d'un réseau est une bonne estimation des probabilités a posteriori d'appartenance aux classes. Cette fonctionnalité sera mise à profit pour localiser la gamme de l'épaisseur

de couches minces, facilitant ainsi sa détermination quantitative.

Les échantillons qui seront étudiés dans ce chapitre concernent un dépôt de résine sur un substrat de verre. Ils sont élaborés par spin coating au Laboratoire Hubert Curien de l'université de Saint Etienne (UMR CNRS 5516). Les indices optiques du substrat sont connus, mesurés par ellipsométrie au laboratoire DIOM (annexe A).

Pour caractériser les couches de résine, le modèle géométrique le plus simple est considéré : il s'agit d'une couche, d'indice N et d'épaisseur h , déposée sur un substrat (figure II.1). Nous supposons par la suite que la couche ne présente aucune rugosité de surface. Toutefois, sa modélisation peut être prise en compte par l'adjonction d'un paramètre supplémentaire.

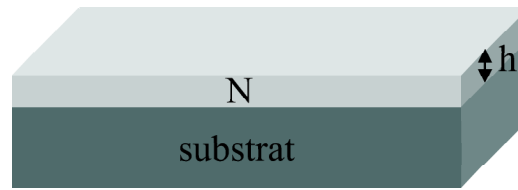


FIG. II.1: Modélisation structurale des échantillons composés d'une couche de résine d'indice N et d'épaisseur h , déposée sur un substrat d'indice connu.

La loi de dispersion $N(\lambda)$ retenue pour la couche de résine est celle de Cauchy (utilisée classiquement pour les diélectriques), donnée par :

$$N(\lambda) = A + \frac{B \cdot 10^4}{\lambda^2} + \frac{C \cdot 10^9}{\lambda^4} \quad (\text{II.1})$$

où λ est exprimée en nm, B en nm^2 , C en nm^4 et A est sans unité.

Remarque : Cette dispersion peut être décrite par d'autres lois plus complexes. Toutefois, dans un souci didactique nous nous limiterons volontairement à la loi de Cauchy parfaitement déterminée par la connaissance des trois paramètres A , B et C . Bien entendu, les résultats présentés par la suite peuvent être étendus au cas plus complexes présentant un nombre de paramètres supérieur à 3.

II.1 Histoire et applications des réseaux de neurones

La reproduction artificielle de l'intelligence humaine constitue le rêve que beaucoup de chercheurs de notre siècle souhaitent réaliser. De nombreux auteurs ont permis de faire évoluer les recherches en neurosciences. La première pierre dans ce domaine date de 1890 où W. James, un célèbre psychologue américain, introduit le concept de mémoire associative. Des travaux multidisciplinaires en physique, en psychologie et en neurophysiologie sont aussi développés par des scientifiques tels Hermann von Helmholtz, Ernst Mach et Ivan Pavlov. A cette époque, il s'agissait de théories plutôt générales sans modèle mathématique précis d'un neurone.

Il a fallu attendre l'année 1943 pour que McCulloch, chercheur en neurologie, et Pitts, mathématicien, proposent le premier modèle mathématique et informatique du neurone

biologique. Il s'agit d'un neurone binaire, c'est-à-dire dont la sortie prend les valeurs 0 ou 1. Ce sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques plus ou moins complexes. C'est ensuite avec D. Hebb, neuropsychologue canadien, qu'arrive une nouvelle vague de développement. Il publia en 1949 son livre intitulé *The Organization of Behavior* [92] dans lequel il expose ses idées sur l'apprentissage pour la première fois. Cet ouvrage sera considéré comme source d'inspiration pour le développement des modèles informatiques d'apprentissage et des systèmes adaptatifs ultérieurs.

Un des succès remarquables de cette discipline remonte à l'année 1957, lorsque F. Rosenblatt, scientifique, invente le Perceptron [93] afin de comprendre la mémoire humaine ainsi que les processus cognitifs. Il construit la première machine basée sur ce modèle et l'applique au domaine de reconnaissance de formes. En 1960, B. Widrow et T. Hoff développent le réseau AdaLiNe (Adaptive Linear Neuron), constitué d'un unique neurone effectuant la combinaison linéaire de ses entrées. Ils proposent la minimisation des erreurs quadratiques en sortie comme algorithme d'apprentissage du réseau.

Vers la fin des années 1960, M. Minsky et S. Papert montrent dans leur livre intitulé *Perceptrons* [94] les limites des réseaux de neurones de type perceptron, notamment l'impossibilité de traiter des problèmes non linéaires. Suite aux idées publiées dans ce livre, la plupart des chercheurs ont abandonné ce domaine avec une chute remarquable des fonds investis.

Heureusement, certains chercheurs ont décidé de ne pas baisser les bras en continuant à développer de nouvelles architectures et algorithmes plus puissants. C'est le cas de T. Kohonen et J. Anderson qui ont mis au point de nouveaux réseaux pouvant servir de mémoires associatives, ainsi que S. Grossberg qui a mis au point ce que l'on appelle aujourd'hui les réseaux auto-organisés. Enfin, l'invention de l'algorithme de rétropropagation de l'erreur dans les années 1980 fut une véritable résurrection du domaine des réseaux de neurones, notamment pour l'apprentissage des perceptrons. Depuis ce temps, cette voie occupe une place importante dans différentes applications grâce au développement continu de nouvelles théories, de nouvelles architectures et de nouveaux algorithmes.

Sans faire un exposé exhaustif des applications des réseaux de neurones, nous en citerons quelques unes dans lesquelles ces outils mathématiques ont montré qu'ils constituent aujourd'hui des alternatives fiables pour la résolution de certains problèmes complexes.

Les réseaux de neurones ont connu sans doute leur envolée dans le domaine de la reconnaissance de formes, des visages humains ou des objets. Le premier modèle dans cette thématique a été développé aux laboratoires AT&T Bell [95] permettant la reconnaissance des codes postaux fournis par les services postaux américains.

En robotique mobile par exemple, les réseaux de neurones peuvent intervenir au niveau du pilotage pour déterminer les actions souhaitables en fonction des consignes de cap et de vitesse [96].

Le marketing est aussi l'un des domaines où les réseaux de neurones sont utiles. Les entreprises les utilisent pour diminuer les tâches administratives de traitement d'énormes bases de données. On appelle cette méthode, le *data mining* (ou déminage de données).

Aujourd'hui, les réseaux de neurones sont présents dans presque tous les domaines de la science, employés pour le traitement d'informations de plus en plus complexes : en bio-ingénierie pour la prédiction des propriétés chimiques de molécules [97] en estimant les relations structure-activité (en anglais QSQR pour *Quantitative Structure-Activity*

Relations), en climatologie pour l'étude de la variabilité du système climatique terrestre et ses impacts [98].

Finalement, les propriétés mathématiques intéressantes des réseaux de neurones ont suscité l'attention de certains chercheurs en optique et microélectronique pour la caractérisation des formes géométriques de structures intégrées. Ce point sera abordé par la suite dans ce travail. La partie suivante détaillera le principe des réseaux de neurones, ainsi que l'architecture qui sera adoptée par la suite dans nos travaux.

II.2 Principe des réseaux de neurones

Par définition, un RN est constitué d'un ou plusieurs neurones. Il nous paraît ainsi judicieux d'expliquer le fonctionnement mathématique d'un simple neurone afin de permettre la compréhension du principe du RN. Nous verrons par la suite les architectures possibles d'un RN selon la façon dont sont regroupés les neurones entre eux. L'architecture qui sera particulièrement détaillée est celle des réseaux à couches. Elle possède des propriétés intéressantes qui seront également exposées.

II.2.1 Du neurone biologique au modèle mathématique

Le cerveau humain est composé de cellules distinctes appelées *neurones* formant un ensemble dense d'environ 10 à 100 milliards d'unités interconnectées.

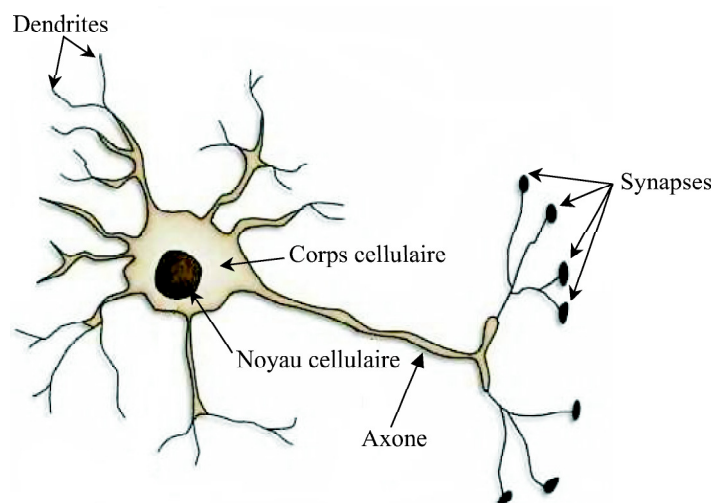


FIG. II.2: Représentation schématique d'un neurone biologique.

L'architecture biologique du neurone est présentée sur la figure II.2. Les *dendrites* sont considérées comme les portes d'entrées des neurones. Elles sont les réceptrices principales du neurone permettant de capter les signaux lui parvenant. Ensuite, un traitement de toutes ces informations a lieu au niveau du *corps cellulaire*. Les principales lignes de transmission du système nerveux sont les *axones*. Ils servent de moyen de transport pour les signaux émis par le neurone, vers d'autres neurones. L'information va transiter de neurone en neurone le long de l'axone jusqu'aux muscles via un potentiel d'action (un

signal électrique). La *synapse* désigne les zones de contact entre neurones, zones spécialisées dans la transmission de l'information. Elles peuvent être inhibitrice ou excitatrice de signal pondérant l'information arrivant au corps cellulaire par les dendrites.

Par analogie avec la biologie, un neurone artificiel (figure II.3) reçoit l'information provenant des entrées x_i ($i = 1, 2, \dots, n$) par l'intermédiaire des connexions dont on affecte à chacune d'elles un poids w (en anglais *weights*) pondérant l'information reçue, appelé aussi *poids synaptique*. Il fonctionne en deux étapes :

La première phase représente le pré-traitement des données reçues en calculant le potentiel v_j du neurone j par la fonction suivante :

$$v_j = b_j + \sum_{i=1}^n w_{j,i} x_i \quad (\text{II.2})$$

où $w_{j,i}$ désigne le poids de la connexion liant le neurone j à l'entrée i , et b_j est un terme constant appelé *biais* qui peut être considéré comme le poids d'une entrée x_0 égale à 1. Ainsi, la relation II.2 s'écrit plus simplement :

$$v_j = \sum_{i=0}^n w_{j,i} x_i \quad (\text{II.3})$$

Remarque : Les neurones que nous allons utiliser sont des neurones linéaires qui effectuent une somme pondérée des entrées. Mais, il en existe d'autres dans la littérature (neurone distance, neurone polynomial, neurone de type noyau, etc...).

Dans la deuxième phase, une *fonction de transfert* g appelée également *fonction d'activation* calcule la valeur de l'état interne s_j du neurone j à partir de la valeur du potentiel v_j . Cette valeur désignera la sortie du neurone :

$$s_j = g(v_j) = g\left(\sum_{i=0}^n w_{j,i} x_i\right) \quad (\text{II.4})$$

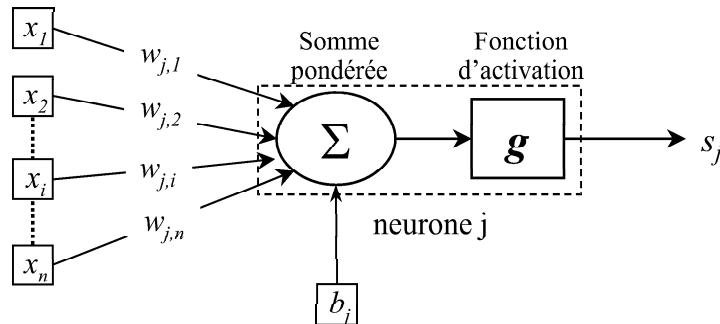


FIG. II.3: Représentation d'un neurone mathématique. Les variables x_i désignent les entrées, la sortie est notée s_j , le biais b_j et les poids synaptiques sont désignés par $w_{j,i}$.

Ainsi, le neurone peut être défini mathématiquement comme étant une fonction algébrique, a priori non linéaire (suivant g) et bornée, des entrées x_i et paramétrée par les

poids $w_{j,i}$:

$$s_j = g(x_i, w_{j,i}) = g_{w_{j,i}}(x_i) \quad (i = 1, 2, \dots, n) \quad (\text{II.5})$$

Plusieurs possibilités existent pour spécifier la nature de la fonction d'activation du modèle. Les fonctions mathématiques les plus souvent utilisées sont :

- La fonction seuil (figure II.4 (a))

$$\begin{aligned} g(x) &= 0 & \text{si} & & x < 0 \\ g(x) &= 1 & \text{si} & & x \geq 0 \end{aligned} \quad (\text{II.6})$$

- La fonction identité (dite linéaire) (figure II.4 (b))

$$g(x) = x \quad \forall x \quad (\text{II.7})$$

- La fonction sigmoïde (figure II.4 (c)), dont les formes les plus utilisées sont la fonction logistique (II.8) et la fonction tangente hyperbolique (II.9) :

$$f(x) = \frac{1}{1 + e^{-Kx}} \quad (\text{II.8})$$

$$f(x) = \frac{e^{Kx} - 1}{e^{Kx} + 1} = \tanh\left(\frac{K}{2}x\right) \quad (\text{II.9})$$

La valeur du paramètre K est généralement prise égale à 1.

- La fonction gaussienne d'écart type unité (figure II.4 (d))

$$f(x) = e^{-x^2} \quad (\text{II.10})$$

Toutes ces fonctions sont continues et différentiables (sauf la fonction seuil). L'une des fonctions les plus souvent utilisées est la fonction sigmoïdale. En effet, elle assure le caractère non linéaire du modèle neuronal qui en fait un approximateur universel comme on le verra plus tard.

II.2.2 Architecture des réseaux de neurones

Comme il a été défini auparavant (équation II.5), un neurone réalise tout simplement une fonction éventuellement non linéaire, paramétrée, de ses entrées. L'association de ces neurones en réseau permet la composition des fonctions réalisées par chacun d'eux, on parle dans ce cas de *réseau de neurones*.

Dans un réseau, les neurones sont connectés les uns aux autres, permettant d'échanger les informations entre eux au moyen des connexions qui les relient. On distingue deux grandes familles de réseaux de neurones : les réseaux *bouclés* et les réseaux *non bouclés*.

Dans un réseau bouclé, appelé aussi réseau récurrent ou dynamique, il est possible de trouver au moins un *cycle* où l'information peut revenir à son point de départ. Cela suppose que la sortie d'un neurone peut être fonction d'elle même. Or cela n'est possible que si la notion de temps est prise en compte. Ce type de réseau n'entrera pas dans le cadre de notre travail et ne sera pas détaillé par la suite. Ce type de réseau est notamment utilisé pour la modélisation dynamique de processus non linéaires [99].

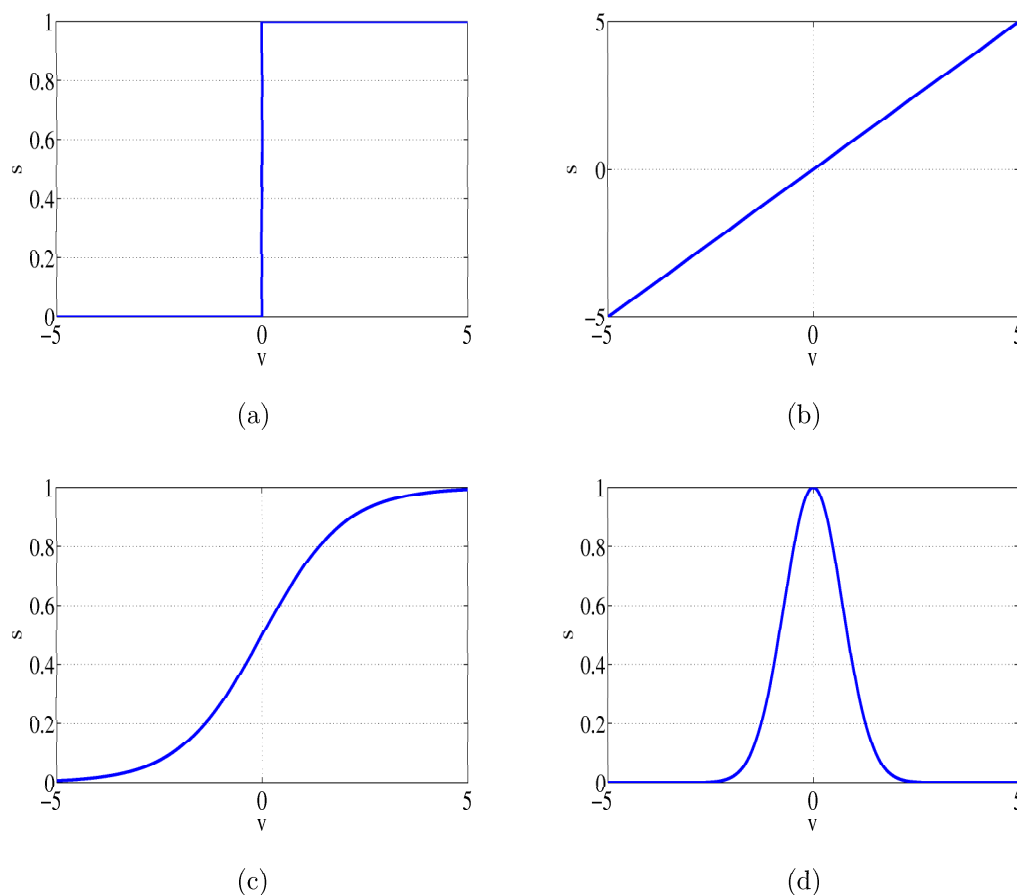


FIG. II.4: Exemples de fonctions de transfert g : (a) seuil, (b) linéaire, (c) sigmoïde (logistique $K=1$) et (d) Gaussienne.

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonctions algébriques de ses entrées par composition des fonctions réalisées par chacun des neurones le composant. L'entrée au réseau désigne l'information reçue par ce dernier de l'environnement extérieur. Dans un tel réseau, l'information circule dans une seule direction, de l'entrée vers la sortie. Ce type de réseaux n'autorise pas le retour en arrière de l'information. Le temps ne joue aucun rôle fonctionnel dans un réseau de neurones non bouclé. Pour cette raison, ces réseaux sont aussi qualifiés de réseaux *statiques*, par opposition aux réseaux dynamiques.

On appelle *neurones de sortie*, les neurones qui effectuent le dernier calcul dans le réseau c'est à dire ceux dont les sorties coïncident avec les sorties du réseau. Ceux qui effectuent les calculs intermédiaires, entre les entrées et les neurones de sortie, sont appelés *neurones cachés*.

Il existe ainsi plusieurs possibilités selon lesquelles les neurones peuvent être connectés pour former un réseau. L'architecture la plus générale est celle du *réseau complètement connecté* (figure II.5).

Dans ce type d'architecture, il faut noter la présence d'une hiérarchie au sein des neurones : chacun d'entre eux est connecté à toutes les entrées x_i du réseau ainsi qu'aux sorties des neurones précédents. Ce type de réseau possède le nombre maximal de connexions que l'on peut avoir.

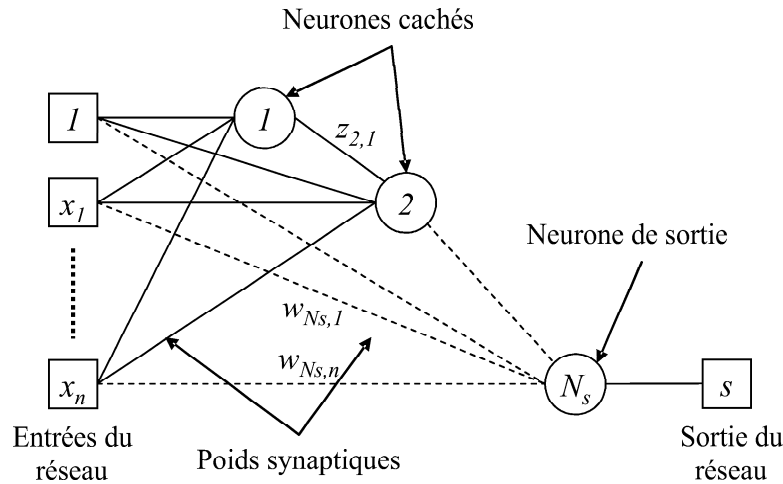


FIG. II.5: Architecture d'un réseau de neurones non bouclé complètement connecté à n entrées $N_s - 1$ neurones cachés et un neurone de sortie. w désigne les connexions entre les entrées et tous les neurones et z ceux entre les neurones.

A partir de ce modèle, il est possible de former de réseaux de moindre complexité en supprimant certaines connexions et en regroupant les neurones; en particulier par couches. Ce type de réseau est probablement le plus utilisé puisqu'il possède des propriétés intéressantes, et fera l'objet du paragraphe suivant.

II.2.3 Cas particulier des perceptrons multicouches

Dans ce paragraphe, nous allons détailler le principe des réseaux à couches en insistant sur les avantages qu'ils possèdent par rapport à d'autres architectures. Nous donnerons également leur représentation mathématique.

II.2.3.1 Architecture et définition

Les réseaux à couches (figure II.6), appelés également Perceptrons Multi-Couches (PMC) sont des réseaux de neurones pour lesquels les neurones sont organisés en couches successives. L'information, provenant des entrées, circule vers les sorties sans retour en arrière et, toute connexion entre neurones d'une même couche est interdite. Un neurone ne peut donc transmettre son état qu'aux neurones appartenant à une couche postérieure à la sienne. La couche recevant l'information à traiter de l'environnement extérieur est appelée couche d'entrée. La couche de sortie est composée d'autant de neurones que le nombre de sorties du réseau. Les couches constituées des neurones effectuant des calculs intermédiaires sont appelées couches cachées.

Choisir ainsi l'architecture d'un PMC consiste à fixer le nombre de couches, le nombre de neurones par couche, la nature des différentes connexions entre les neurones et la fonction d'activation des neurones dans chaque couche.

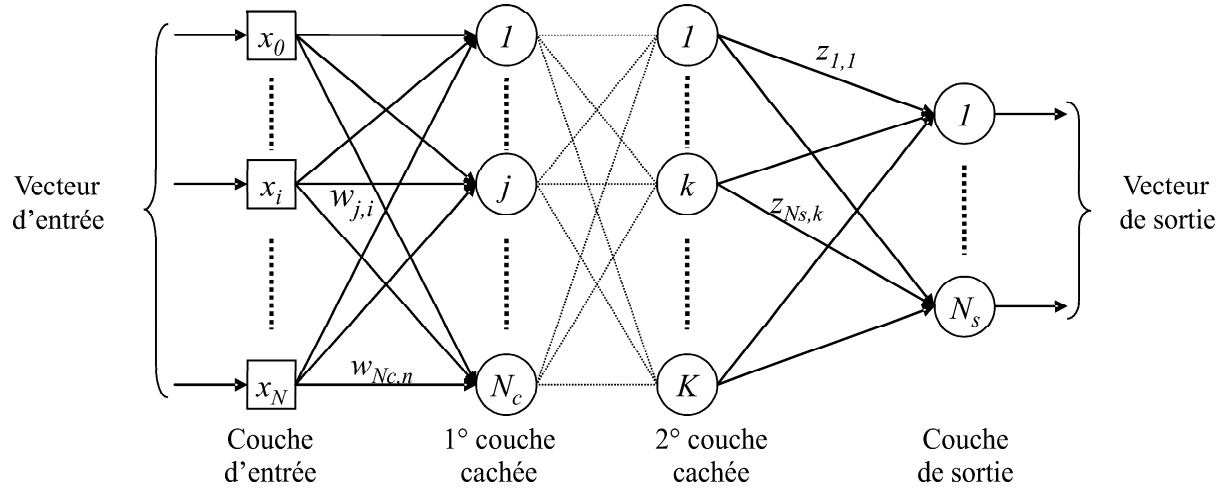


FIG. II.6: Structure d'un réseau de neurones de type PMC à deux couches cachées.

II.2.3.2 Propriété d'approximation universelle

La propriété d'approximation universelle a été démontrée par Cybenko [100] et Funahashi [101] et peut s'énoncer de la façon suivante :

Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire.

Remarque : Cette propriété s'applique pour les réseaux présentés au paragraphe précédent, mais aussi pour les fonctions ondelettes [102] et radiales où la sortie d'un neurone j peut s'écrire :

$$s_j = \exp \left[\frac{\sum_{i=1}^n (x_i - w_{j,i})^2}{2w_{j,0}^2} \right] \quad (\text{II.11})$$

avec w_i sont les coordonnées du centre de la gaussienne et $w_{j,0}$ son écart type.

Cette propriété justifie notre choix des réseaux à une seule couche cachée dans la suite de notre travail. Le seul degré de liberté qui subsiste pour déterminer l'architecture du réseau est le nombre de neurones cachés. Ce nombre doit être choisi convenablement selon la nature du problème pour obtenir la précision souhaitée.

II.2.3.3 Propriété de parcimonie

Lors de la modélisation d'un processus physique quelconque à partir de données, on cherche toujours un nombre minimal de paramètres ajustables, notés \vec{a} précédemment (partie I.3.1), permettant d'obtenir les résultats les plus satisfaisants possibles. Dans cette optique, Hornik [103] a montré que :

Si le résultat de l'approximation (c'est-à-dire la sortie du réseau de neurones) est une fonction non linéaire des paramètres ajustables, elle est plus parcimonieuse que si elle est une fonction linéaire de ces paramètres. De plus, pour des réseaux de neurones à fonction d'activation sigmoïdale, l'erreur commise dans l'approximation varie comme l'inverse du nombre de neurones cachés, et elle est indépendante du nombre de variables de la fonction

à approcher. Par conséquent, pour une précision donnée, donc pour un nombre de neurones cachés donné, le nombre de paramètres du réseau est proportionnel au nombre de variables de la fonction à approcher.

Dans le cas des réseaux de neurones à fonctions d'activation sigmoïdale, les sorties ne sont pas linéaires par rapport aux paramètres ajustables (les poids synaptiques $w_{j,i}$) et la parcimonie peut être parfaitement appliquée.

II.2.3.4 Formulation mathématique

En pratique, il est rare de rencontrer des problèmes où l'on doit approcher une fonction connue. Dans la plupart des cas, on cherche à modéliser un processus physique à partir de mesures expérimentales. Ainsi, considérons que l'on dispose d'un ensemble de n_t observations mesurables $\vec{y}^q = [y_1^q, y_2^q, \dots, y_m^q, \dots, y_{N_s}^q]$ ($q = 1 \dots n_t$) supposées liées aux paramètres physiques $\vec{x}^q = [x_1^q, x_2^q, \dots, x_i^q, \dots, x_N^q]$. Le but est de déterminer une forme mathématique de la relation univoque liant \vec{x} à \vec{y} , valable dans le domaine où les mesures ont été effectuées sachant que ces dernières sont en nombre fini et entachées de bruit de mesure. Les données disponibles se présentent donc sous la forme de couple $\{\vec{x}^q, \vec{y}^q\}$.

Dans le cas où le PMC (à 1 couche cachée) possède n entrées par exemple et la couche de sortie contient N_s sorties, on peut considérer que la sortie m du RN réalise la fonction $F(W, Z, \vec{x})$.

$$s_m(\vec{x}, W, Z) = \sum_{j=0}^{N_c} z_{m,j} \left[g \left(\sum_{i=0}^N w_{j,i} x_i \right) \right] \quad (\text{II.12})$$

avec W la matrice poids des connexions entre la couche cachée et la couche d'entrée et Z celle des connexions entre la couche de sortie et la couche cachée, exprimées respectivement par :

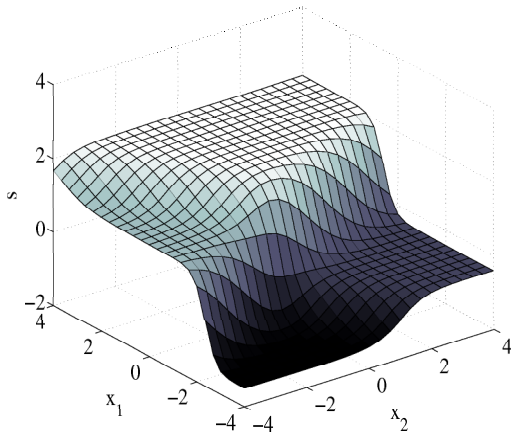
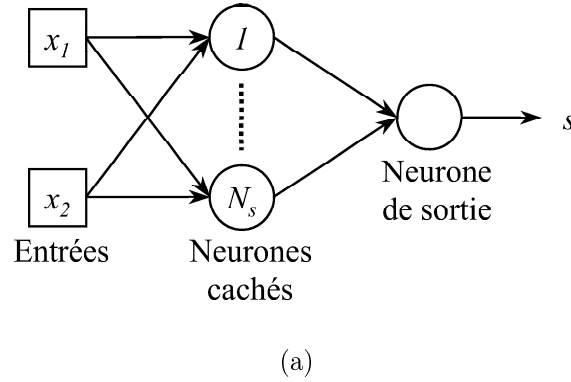
$$W = \begin{bmatrix} w_{1,1} & \dots & w_{1,i} & \dots & w_{1,N} \\ \vdots & \ddots & & & \vdots \\ w_{j,1} & & w_{j,i} & & w_{j,N} \\ \vdots & & & \ddots & \vdots \\ w_{N_c,1} & \dots & w_{N_c,i} & \dots & w_{N_c,N} \end{bmatrix}; \quad Z = \begin{bmatrix} z_{1,1} & \dots & z_{1,j} & \dots & z_{1,N_c} \\ \vdots & \ddots & & & \vdots \\ z_{m,1} & & z_{m,j} & & z_{m,N_c} \\ \vdots & & & \ddots & \vdots \\ z_{N_s,1} & \dots & z_{N_s,j} & \dots & z_{N_s,N_c} \end{bmatrix} \quad (\text{II.13})$$

Rappelons que $w_{j,i}$ désigne le poids de la liaison liant une entrée x_i au neurone j de la couche cachée et, $z_{m,j}$ désigne le poids de la liaison entre la sortie m et le neurone j de la couche cachée.

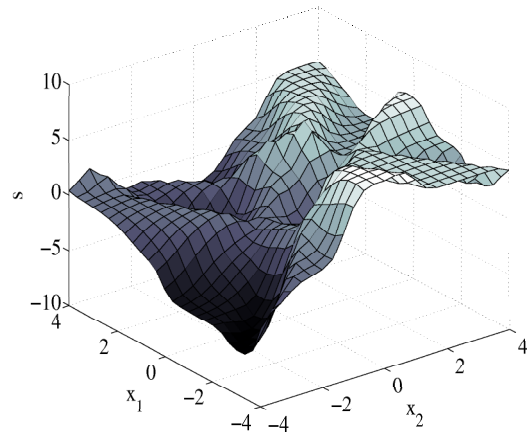
Il est possible de formuler la réponse d'un RN par une approche statistique. Nous allons considérer que les valeurs \vec{y}^q mesurées pour une observation q donnée ne sont pas des valeurs vraies mais présentent une certaine variabilité qui peut provenir notamment de la précision de l'appareil de mesure. De façon à prendre en compte ces incertitudes, il est courant de faire l'hypothèse que chaque mesure y_m^q est une réalisation d'une variable aléatoire Y_m^q supposée de densité de probabilité gaussienne d'écart type σ_m . Ainsi, si on considère que ces σ_m sont indépendants, la vraie valeur de y_m^q correspondant à \vec{x}^q est approchée par l'espérance mathématique de Y_m^q sachant \vec{x}^q , notée $E[Y_m^q/\vec{x}^q]$. Le but du

RN est ainsi de déterminer pour chacune de ses sorties :

$$s_m(\vec{x}^q, W, Z) = E[Y_m^q / \vec{x}^q] \quad (\text{II.14})$$



(b) 2 neurones cachés



(c) 15 neurones cachés

FIG. II.7: Sortie s calculée par un (a) PMC à deux entrées x_1 et x_2 variant dans l'intervalle $[-4; 4]$, à une couche cachée (fonction d'activation sigmoïdale) et à une sortie (fonction d'activation linéaire). Les figures (b) et (c) correspondent respectivement à s pour $N_c = 2$ et $N_c = 15$, lorsque les poids sont fixés.

A titre d'exemple, nous avons tracé sur les figures II.7(b) et II.7(c) les différentes fonctions réalisées par l'unique sortie d'un PMC (figure II.7(a)) à 1 couche cachée pour 2 entrées (x_1, x_2).

La fonction d'activation de la couche cachée est sigmoïdale (équation II.9), celle de la couche de sortie est linéaire afin d'assurer les propriétés d'approximation parcimonieuse (voir paragraphe II.2.3.2 et II.2.3.3). Les entrées, x_1 et x_2 , peuvent varier toutes deux dans l'intervalle $[-4; 4]$. Les poids entre la couche d'entrée et la couche cachée sont choisis aléatoirement dans l'intervalle $[-4; 4]$, ceux entre la couche cachée et la couche de sortie dans l'intervalle $[-2; 2]$. Le choix de ces intervalles est totalement indépendant et sert juste d'une illustration pour l'exemple considéré.

La figure II.7(b) représente le cas où $N_c = 2$ alors que la figure II.7(c) correspond à $N_c = 15$. Les fonctions réalisées par un réseau possédant un nombre élevé de neurones cachés sont beaucoup plus complexes que celles réalisées par un nombre de neurones plus

restreint. En effet, la complexité de la fonction réalisée est directement reliée au nombre et à la qualité des calculs intermédiaires réalisés par les neurones cachés. En fait, plus N_c est élevé plus le nombre de paramètres (de poids) est élevé et plus la fonction possède de degrés de liberté.

Nous venons de présenter le principe général des réseaux de neurones en détaillant plus particulièrement l'architecture ainsi et les propriétés des PMC à une couche cachée, qui sont considérés comme de véritables approximateurs parcimonieux. Intéressons nous à présent à la mise en œuvre de telles structures.

II.3 Fonctionnement des réseaux de neurones

L'utilisation d'un RN nécessite deux étapes distinctes : l'entraînement et l'exploitation. La première étape est primordiale pour la mise en place d'un RN. Nous allons débiter cette partie par définir l'apprentissage. Nous détaillerons ensuite les différentes étapes de traitement qui doivent être employées afin d'assurer un fonctionnement correct.

II.3.1 Phase d'apprentissage

L'apprentissage permet de trouver les poids convenables permettant d'établir la relation entre les entrées et les sorties du RN, selon le but recherché. Dans ce paragraphe, nous allons définir le mode d'apprentissage utilisé et l'algorithme choisi pour achever cette tâche.

II.3.1.1 Définition et type d'apprentissage

L'apprentissage peut être classé en deux catégories : l'apprentissage *supervisé* et l'apprentissage *non supervisé*.

Supposons que l'on doive modéliser un processus physique, liant le vecteur d'entrée \vec{x} au vecteur de sortie \vec{y} . Nous avons vu, qu'en choisissant une architecture adéquate de RN, ce dernier peut générer une famille de fonctions paramétrées par les valeurs de ses poids synaptiques W et Z (équation II.12). L'apprentissage consiste donc à modifier l'ensemble des poids W et Z afin de trouver parmi toutes ces fonctions possibles celles qui permettent aux sorties \vec{s} du RN d'approcher au mieux les observations \vec{y} par la seule connaissance de \vec{x} . D'une manière générale, un algorithme d'apprentissage permet de modifier les poids de façon itérative. Le poids $w_{j,i}$ à l'itération $k + 1$ peut ainsi s'écrire :

$$w_{j,i}^{k+1} = w_{j,i}^k + \Delta w_{j,i} \quad (\text{II.15})$$

tel que $\Delta w_{j,i}$ est la modification apportée.

On distingue l'*entraînement adaptatif* si les poids sont modifiés après chaque passage de couple du corpus. Dans le cas où ils sont modifiés après le passage de l'ensemble du corpus, l'entraînement est dit *non adaptatif* (*batch training* en anglais), et sera utilisé par la suite.

Dans l'apprentissage supervisé (ou analogie du maître-élève), l'utilisateur fournit les valeurs désirées à la sortie pour chaque entrée connue. Les poids sont ainsi modifiés afin

de minimiser l'écart entre les sorties \vec{s} calculées par le RN et les sorties désirées connues \vec{y} . La fonction coût la plus utilisée, et dont nous nous sommes servis pour la mise en œuvre de notre RN, est l'erreur quadratique notée C . Elle consiste à calculer la somme des carrés des erreurs entre les vecteurs \vec{s} et \vec{y} sur l'ensemble des n_t données utilisées pour l'apprentissage sur chacune des N_s sorties :

$$C(W, Z) = \sum_{q=1}^{n_t} \left[\sum_{m=1}^{N_s} (y_m^q - s_m^q)^2 \right] \quad (\text{II.16})$$

Ainsi, il est nécessaire de disposer d'une base de données appelée *corpus d'apprentissage*, contenant n_t couples d'exemples $\{\vec{x}^q, \vec{y}^q\}$. Dans l'apprentissage non supervisé, on ne possède pas d'information a priori sur les sorties du RN. Il n'existe pas dans ce cas un superviseur ou un professeur pour fournir au RN des exemples connus. Ce type d'apprentissage ne sera pas détaillé car il concerne des RN particuliers dédiés principalement à la fouille de données.

II.3.1.2 Algorithme de rétropropagation de l'erreur

L'apprentissage du RN n'est autre qu'un problème d'optimisation paramétrique, dont le but est de trouver les poids synaptiques adéquats qui minimisent la fonction C (équation II.16). Il peut être achevée par l'emploi d'une des méthodes d'optimisation classiques présentées au paragraphe I.3.3. Or, la plupart de ces méthodes nécessitent la connaissance à chaque itération des dérivées premières et secondes de la fonction coût par rapport au poids du RN. Dans le cas des perceptrons multicouches, le problème a longtemps résidé dans le calcul de ces dérivées dans la (ou les) couches cachées. En effet, il était difficile de quantifier la sensibilité des neurones intermédiaires sur les sorties du RN.

Pour résoudre ce problème, Rumelhart proposait un algorithme original appelé algorithme de *rétropropagation du gradient* [104] (Annexe A) permettant de calculer de manière récursive le gradient de la fonction coût relatif aux paramètres des couches cachées, à savoir $\partial C / \partial w_{j,i}$.

Cette avancée a permis d'appliquer n'importe quelle méthode d'optimisation utilisant ce gradient pour la phase d'apprentissage du RN, ce qui a joué un rôle fondamental dans le développement et l'application de ces structures. Pour notre cas, nous utiliserons l'algorithme de Levenberg-Marquardt (voir I.3.3).

En général, la surface d'erreur à optimiser est remplie de minima locaux en raison du très grand nombre de paramètres à ajuster. C'est pour cela qu'il faut relancer la procédure d'apprentissage plusieurs fois. Il existe dans la littérature des études permettant un entraînement utilisant les algorithmes génétiques [105, 106]. Cette méthode d'apprentissage n'est pas assez documentée à l'heure actuelle et ne sera pas utilisée dans notre travail et les potentialités d'un tel couplage ne seront pas abordées dans ce manuscrit.

II.3.2 Traitement des données d'entraînement

En général, les données d'entraînement du RN sont de nature physique. Le traitement de ces données est essentiel pour favoriser le bon déroulement de l'apprentissage. Nous indiquerons dans les paragraphes suivants les différentes étapes nécessaires afin de faciliter la phase d'apprentissage.

II.3.2.1 Initialisation des paramètres internes

Comme on vient de le voir, l'apprentissage dépend fortement des conditions initiales. Quelle que soit la méthode employée pour effectuer l'apprentissage, il est nécessaire au début d'initialiser les valeurs des poids W et Z du RN.

Il faut en général s'assurer qu'en début d'apprentissage, les valeurs des sorties des neurones ne sont pas trop voisines de -1 ou $+1$. En effet, le gradient de la fonction coût, qui permet de guider le processus de minimisation, dépend de la dérivée de la fonction d'activation des neurones cachés par rapport au potentiel. Or, les dérivées des sigmoïdes sont voisines de zéro au voisinage de leurs saturations (voir figure II.4). Dans cette situation, le neurone correspondant ne contribuera que très peu au processus d'apprentissage.

La première procédure d'initialisation consiste à tirer aléatoirement les poids dans une distribution normale centrée de variance $1/(n_t - 1)$. Si l'on souhaite une distribution uniforme entre $-w_{max}$ et $+w_{max}$, il faut choisir $w_{max} = \sqrt{3(n_t - 1)}$. Dans tous les cas, il faut s'assurer que les biais soient initialisés à zéro.

Il existe également une autre méthode très employée dans le cas des fonctions de transfert sigmoïdales : la méthode de Nguyen et Widrow [107] initialise les poids de sorte que la région active du neurone (à savoir la partie linéaire de la sigmoïde) soit distribuée à peu près de manière égale sur l'espace des entrées.

II.3.2.2 Normalisation des données d'entraînement

Le vecteur d'entrées \vec{x} présente parfois des éléments d'ordres de grandeur et de nature différents. Il est préférable dans ce cas d'effectuer un pré-traitement afin qu'ils aient toutes la même influence sur la conception du modèle.

En pratique, le recours à des données réduites et centrées est recommandé. Dans notre travail, un changement de variable sera effectué avant la phase d'apprentissage du RN. Pour une observation quelconque q , ce changement se traduit par la transformation du vecteur $\vec{x}^q = [x_1^q, x_2^q, \dots, x_i^q, \dots, x_N^q]$ en $\vec{x}'^q = [x_1'^q, x_2'^q, \dots, x_N'^q]$ tel que :

$$x_i'^q = \frac{x_i^q - \mu_i}{\sigma_i} \quad (\text{II.17})$$

avec μ_i et σ_i sont la moyenne et l'écart type calculés sur l'ensemble des données du corpus d'apprentissage correspondant à l'entrée i .

De cette façon, la moyenne de chaque entrée est proche de 0 et son écart type de 1. Les sorties sont également normalisées selon la même technique.

II.3.2.3 Réduction de la dimension de l'espace des variables

Le choix du vecteur d'entrée n'est pas à prendre à la légère. En effet, il est souhaitable qu'il soit le plus petit possible dans le but avoué de favoriser l'apprentissage. Toutefois, toute entrée soupçonnée influencer les sorties \vec{y} doit être prise en compte. Il est ainsi primordial de considérer l'ensemble des informations (physiques ou autres) disponibles pour définir le nombre optimal d'entrées. Si ce choix est difficile ou si le nombre d'entrées est exagérément grand, il est possible d'effectuer une réduction de l'espace des entrées dans une dimension raisonnable pour que l'information soit efficacement exploitée, et surtout pour ne pas augmenter inutilement la complexité du modèle neuronal.

La technique d'*analyse en composantes principales* (ACP) [108] est une des plus anciennes techniques d'analyse statistique, adaptée à l'analyse des données multidimensionnelles. Elle est souvent utilisée pour simplifier la construction d'un modèle lorsque les composantes du vecteur d'entrée sont nombreuses ou lorsqu'elles ne sont pas mutuellement indépendantes. Elle fournit une représentation de l'ensemble de données dans un sous-espace de dimension plus faible. Ainsi, elle permet de réduire le nombre d'entrées, dans lesquels la répartition statistique des données est conservée au mieux. Elle se présente comme une méthode de projection linéaire qui maximise l'inertie du nuage de points des données décrit dans les différents espaces engendrés.

Il faut noter que l'ACP n'aboutit pas à la suppression de certaines entrées, mais à diminuer seulement leur nombre. En effet, chaque nouvelle entrée est une combinaison linéaire des N entrées initiales (figure II.8). Ainsi, d'un point de vue expérimental la mesure de toutes les entrées reste nécessaire.

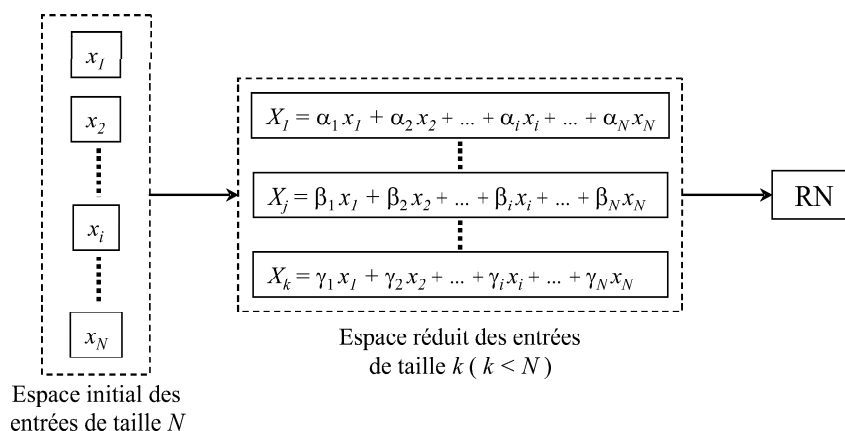


FIG. II.8: Représentation schématique du principe de l'ACP : Réduction de l'espace initial de dimension N en un espace de dimension k tel que ($k < N$), en conservant toutes les données initiales.

La réduction du vecteur d'entrée permet d'atténuer la complexité du RN et donc diminuer le nombre de paramètres à optimiser pendant l'apprentissage. Ceci permet de réduire la complexité de la surface d'erreur lors de la phase d'apprentissage et donc à restreindre le risque permanent de convergence prématurée.

On peut citer également d'autres méthodes comme l'*analyse en composantes curvilinear* [109, 110] qui peut être vue comme une extension non linéaire de l'ACP, et les RN auto-organisés de Kohonen [111] utilisés dans les analyses non linéaires de données. Cependant, ce type de transformation sur les entrées conduit inévitablement à une perte d'information (plus ou moins grande suivant la nature de la projection) qui peut être néfaste.

Cette technique ne sera pas utilisée par la suite, mais nous proposerons une autre méthode plus efficace qui permet de supprimer les entrées moins influentes. Ceci conduira à simplifier l'architecture du RN ainsi que le temps de mesure nécessaire.

II.3.3 Généralisation et Estimation des performances

Un RN entraîné ne peut pas être exploité sans avoir vérifié certains critères. Nous allons à présent indiquer la façon qui nous permettra d'évaluer les performances du RN ainsi

que les principaux critères à prendre en compte afin de juger la qualité de l'entraînement effectué.

II.3.3.1 Dilemme biais-variance

Nous avons vu précédemment qu'il était possible de diminuer fortement la fonction coût sur l'ensemble des observations en mettant suffisamment de neurones cachés lors de la phase d'apprentissage. Or, dans la pratique l'objectif d'une modélisation statistique n'est pas d'ajuster finement un modèle sur un ensemble d'apprentissage, mais d'approcher le modèle qui a servi pour générer ces données. D'après la formulation probabiliste du paragraphe (II.2.3.4), c'est l'espérance mathématique de la variable aléatoire Y_m^q sachant \tilde{x}^q qui est recherché et non une réalisation expérimentale particulière y_m^q .

Ainsi, si le RN est ajustée trop finement à l'ensemble de données, on dit qu'il y a apprentissage par coeur des particularités de cet ensemble au détriment du modèle sous-jacent et le RN est dans ce cas surajusté.

Le compromis biais-variance (ou *dilemme biais-variance*) [112] est un problème très importante en modélisation des données. Le biais rend compte de la différence moyenne entre le modèle et l'espérance mathématique de la grandeur à modéliser, il est donc lié à la valeur du bruit du processus que l'on cherche à modéliser (à savoir σ_m de la formulation probabiliste). Quant à la variance, elle rend compte des différences entre les modèles issus de différentes bases d'apprentissage et donc reflète l'influence du choix de celle-ci.

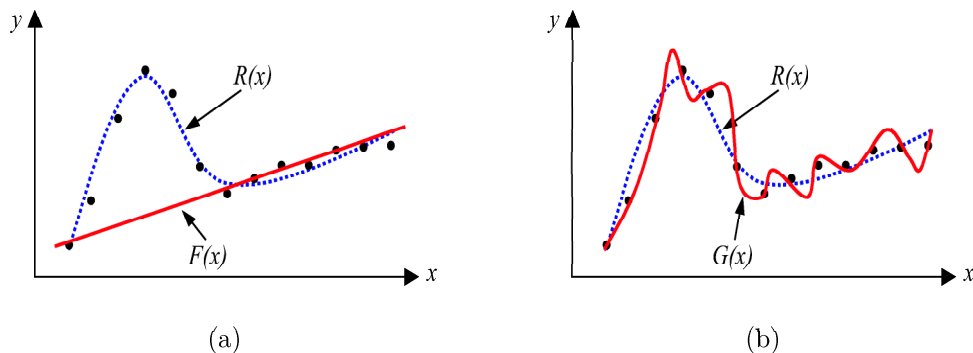


FIG. II.9: Approximation d'une fonction $R(x)$ à partir des données bruitées de celle-ci (points) (a) par un modèle linéaire simple $F(x)$, (b) par un modèle polynomial plus complexe $G(x)$.

Ce compromis peut être expliqué d'une façon simple : considérons le cas d'une fonction $R(x)$ à approximer à partir de mesures bruitées. Si le modèle utilisé pour ajuster cette fonction est simple, par exemple linéaire (figure II.9(a)), l'écart avec le modèle idéal est important. Dans ce cas, le modèle possède un biais très élevé mais sa variance est faible puisqu'il est peu influencé par un jeu de données. En revanche, si le modèle est complexe (par exemple polynomial), c'est à dire possédant plus de degrés de liberté (figure II.9(b)), il sera capable de s'ajuster au plus près des données. On dit dans ce cas que son biais est faible mais sa variance est élevée. Le modèle est dans ce cas trop sensible au choix des données utilisées. C'est le cas du *surajustement*.

Ainsi, un RN entraîné correctement doit être capable de répondre convenablement à l'introduction d'une nouvelle entrée inconnue. On parle de capacité de généralisation. Notons que cette nouvelle entrée doit forcément être incluse dans le domaine d'entraînement usité.

II.3.3.2 Estimation des performances globales du réseau

Une fois le RN entraîné, l'étape qui suit est destinée à estimer ses performances globales. Pour cela, nous allons utiliser une partie de la base de données appelée corpus test, composé de n_{test} couples exemples. En effet, il est important d'utiliser des données n'ayant pas servi directement ou indirectement lors de la phase d'apprentissage pour éviter tout biais dans l'estimation des performances.

Les performances globales du RN sont obtenues par le calcul de l'erreur quadratique moyenne de la même façon que l'équation II.16 mais évaluée sur les n_{test} observations. Ainsi, l'erreur E_{test}^m sur la sortie s_m est donnée par :

$$E_{test}^m = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (y_m^i - s_m^i)^2} \quad (\text{II.18})$$

Cette erreur constitue une première estimation de la qualité de l'apprentissage réalisé. Cependant, elle doit être comparée à l'erreur quadratique moyenne calculée sur l'ensemble d'entraînement E_{entr}^m donnée par la formule analogue :

$$E_{entr}^m = \sqrt{\frac{1}{n_{entr}} \sum_{i=1}^{n_{entr}} (y_m^i - s_m^i)^2} \quad (\text{II.19})$$

où n_{entr} est le nombre d'exemples contenus dans le corpus dédié exclusivement à la phase d'entraînement.

Cela constitue une seconde vérification pour prévenir le surapprentissage. En effet, dans le cas où E_{entr}^m est très faible et E_{test}^m très élevé, cela signifie que le RN ne généralise pas dans de bonnes conditions. Il a parfaitement appris sur les données d'entraînement mais est défaillant sur des nouvelles données. Ainsi, le meilleur modèle est celui qui possède des erreurs E_{entr}^m et E_{test}^m les plus petites possibles et ayant le même ordre de grandeur.

Les performances globales du RN peuvent être également estimées en calculant l'erreur quadratique moyenne sur tout l'ensemble de N_s sorties :

$$E_{test} = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left[\frac{1}{N_s} \sum_{m=1}^{N_s} (y_m^i - s_m^i)^2 \right]} \quad (\text{II.20})$$

L'erreur quadratique E_{test} estime la performance globale du modèle en terme de généralisation. Elle est utile pour pouvoir comparer rapidement plusieurs RN entre eux. Il faut noter que cette grandeur ne représente pas exactement l'intervalle de confiance associé à l'une des sorties du RN, mais donne l'erreur moyenne réalisé par le RN pour chacune de ses m sorties. En effet, c'est une estimation globale de la différence moyennée entre les observations et les résultats obtenus par le RN. Elle est indépendante du vecteur d'entrée, c'est à dire de la mesure.

En métrologie, une estimation de paramètres donnée sans intervalle de confiance n'a que peu de sens. Il s'agit en fait de la précision avec laquelle le modèle est capable d'ajuster localement sa sortie, pour chaque vecteur d'entrée. Nous détaillerons plus loin dans le manuscrit, la méthode que nous utiliserons pour estimer les barres d'incertitude sur les sorties du RN.

II.3.3.3 Méthodes pour limiter le surajustement

Il existe deux grands types de méthodes pour limiter le problème de surajustement :

- *Les méthodes passives* détectent le surajustement *a posteriori*, c'est à dire après l'entraînement, pour supprimer les mauvais modèles. Parmi les méthodes les plus souvent employées dans cette famille, on compte la validation croisée [113]. On trouve aussi celle qui supprime les modèles dont la matrice jacobienne ($\partial C / \partial w_{j,i}$) n'est pas de rang plein [114] et une méthode locale appelée LOCL (pour *Local Overfitting Control via Laverages*) [115] qui permet d'estimer l'influence de chaque exemple sur le modèle.
- *Les méthodes actives* interviennent au cours de la phase d'apprentissage pour empêcher le modèle de faire un surajustement. On peut citer les techniques de régularisation comme l'arrêt prématuré ou la pénalisation de la fonction coût. Ces méthodes ne cherchent pas à limiter la complexité du RN, mais elles contrôlent la valeur des poids pendant l'apprentissage. Dans ce contexte, Barlett a montré [116] que la valeur des poids emportait sur leur nombre afin d'obtenir des modèles qui ne sont pas surajustés.

La validation croisée. Le principe de cette méthode consiste à estimer les performances du RN à partir d'exemples n'ayant pas servi pendant l'apprentissage. On divise d'abord la base d'apprentissage en D parties de taille égale (P_1, P_2, \dots, P_D) (figure II.10). On réalise ensuite D apprentissages successifs du RN, en laissant à chaque fois une fraction à côté pour la validation. Sur cette fraction d'exemples, on calcule l'erreur commise par le modèle. A la fin, la performance globale du modèle appelée *score de validation croisée* est calculée en réalisant la moyenne des D erreurs. Le modèle optimal est alors celui qui présente le score de validation le plus faible.

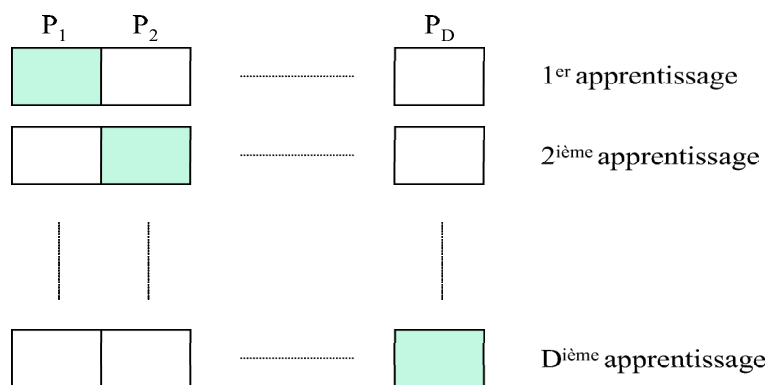


FIG. II.10: Principe de la validation croisée ; les parties grises sont utilisées pour la validation après chaque apprentissage.

Un cas particulier de cette méthode, appelé également méthode de *leave one out* [117], est celui où le nombre D est égal au nombre d'exemples utilisés dans la base d'apprentissage. Dans ce cas, l'erreur est calculée sur un seul exemple après chaque apprentissage.

Le temps de calcul présente le principal inconvénient de cette méthode. Pour le même corpus d'apprentissage, ce temps est d'autant plus grand que le nombre D est élevé. Il est donc maximal dans le cas de la méthode de *leave one out*. D'un autre côté, on ne sait pas a

priori si le nombre d'exemples utilisés est suffisant ou pas pour bien approcher le modèle. Si ce nombre est assez grand, la méthode est assez efficace et donne de bons résultats même avec un petit nombre D . Dans le cas contraire, on doit augmenter le nombre D en gardant un nombre suffisant d'exemples pour les D apprentissages.

Arrêt prématuré. En général, l'apprentissage est arrêté soit lorsqu'un nombre d'itérations est atteint soit lorsque la fonction coût est rendue minimale. Le principe de l'arrêt prématuré, comme son nom l'indique, consiste à arrêter l'apprentissage prématurément, c'est à dire avant que l'algorithme s'ajuste trop finement aux données d'apprentissage. Le surajustement est ainsi évité. Il reste à déterminer le moment où l'on doit arrêter l'apprentissage.

La méthode la plus utilisée consiste à diviser en deux parties l'ensemble des données disponibles. Le premier ensemble, appelé corpus d'apprentissage et souvent le plus grand des deux, sert à l'apprentissage proprement dit du RN. Le deuxième, appelé corpus de validation, contrôle l'arrêt de l'apprentissage. En effet, à chaque itération l'erreur quadratique commise par le RN est calculée sur cet ensemble et noté E_{val} .

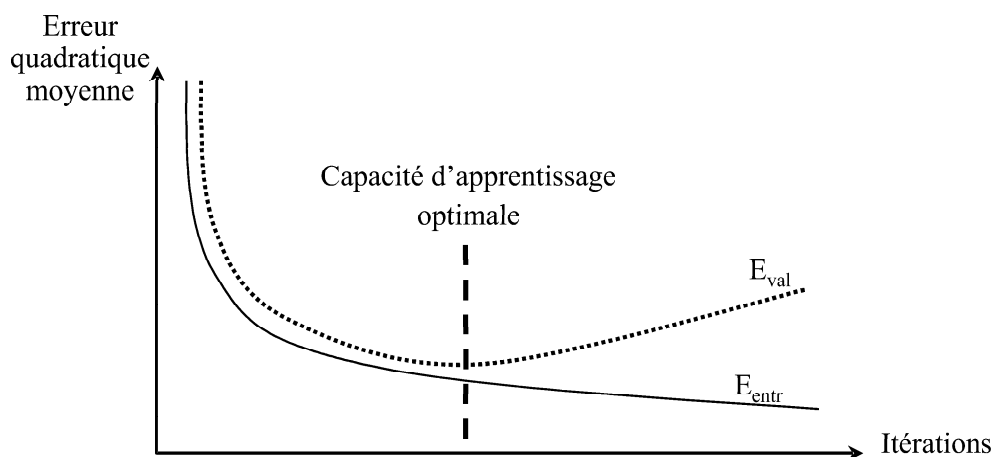


FIG. II.11: Evolution schématique des performances calculées sur les bases des données d'apprentissage E_{entr} et de validation E_{val} .

Au cours des premières itérations, les erreurs E_{entr} et E_{val} tendent à diminuer. L'apprentissage est ensuite arrêté dès que l'erreur de validation atteint un minimum (figure II.11). Dans ce cas, la généralisation du RN est optimale. Si l'entraînement est poursuivi, l'erreur sur la base d'apprentissage continue à diminuer pour un nombre de neurones fixé (voir paragraphe II.2.3.3). Dans la zone où l'erreur de validation commence à s'écarter de l'erreur commise par le RN sur le corpus d'apprentissage, on dit que le RN est surajusté.

Malgré les résultats satisfaisants en pratique, cette méthode conduit souvent à des RN surdimensionnés avec un grand nombre de poids et des neurones inutiles.

Une fois l'entraînement arrêté, il est conseillé de tester les performances du RN sur un autre ensemble de données, appelée corpus test, n'ayant pas servi pendant le processus d'apprentissage afin d'éviter tout biais éventuellement introduit par les données de validation qui contribuent indirectement à l'apprentissage.

Cette méthode est rapide et facile à mettre en œuvre. Il est toutefois difficile de choisir, qualitativement et quantitativement, comment répartir les exemples disponibles

en apprentissage et en validation. Pour tous les RN entraînés par la suite, l'apprentissage sera effectué sur 70% des données disponibles, la validation sur 20%, et les 10% restants formeront le corpus test.

Pénalisation de la fonction coût. Lorsque le potentiel des neurones à fonction d'activation de type sigmoïdale est élevé, la réponse du neurone atteint rapidement la saturation (paragraphe II.3.2.1). Il est ainsi préférable que les sigmoïdes restent au voisinage de leurs zones linéaires, ce qui implique que les poids ne prennent pas des valeurs élevées.

La méthode de régularisation par pénalisation de la fonction coût, dont le principe est d'introduire des contraintes sur la fonction coût, empêche les poids de prendre des valeurs trop élevées. La fonction coût utilisée par cette méthode est alors donnée par :

$$C'(C, W, Z) = C(W, Z) + \alpha \left(\sum_{j=1}^{N_c} \sum_{i=0}^N w_{j,i}^2 + \sum_{m=1}^{N_s} \sum_{j=1}^{N_c} w_{m,j}^2 \right) \quad (\text{II.21})$$

où α est un hyperparamètre qui pondère l'importance des valeurs des poids sur l'erreur C .

Cette méthode est également nommée *weight decay*. On la trouve également dans le cas des modèles de régression linéaires par rapport aux paramètres, appelée *ridge regression* [118].

La difficulté de cette méthode réside dans la détermination du paramètre de régularisation α . Les valeurs des poids vont décroître rapidement pour un choix de α trop grand et le modèle risque d'avoir un biais élevé. Inversement, si α est trop faible, le terme de régularisation a très peu d'effet sur l'apprentissage ce qui conduit à un modèle possédant une variance élevée et un biais très faible. Le surajustement risque ainsi d'apparaître dans ce cas. Une solution pour la détermination de cet hyperparamètre en utilisant une approche bayésienne a été proposé par McKay [119]. Il prend en compte le caractère différent des poids en fonction des couches en considérant plusieurs hyperparamètres.

Nous avons montré dans ce paragraphe que la phase d'apprentissage doit être menée avec attention. En effet, le phénomène de surapprentissage doit être évité pour assurer de bonnes performances en généralisation. Les performances du RN peuvent ensuite être estimées en première approche par le calcul de l'erreur quadratique moyenne sur un ensemble neutre n'ayant pas participé à l'entraînement. Une fois ces considérations validées, le RN peut être utilisé efficacement pour une tâche dédiée.

II.4 Résolution du problème inverse par réseau de neurones

Tout au long de cette présentation du principe de fonctionnement d'un RN, nous avons supposé que le PMC était utilisé pour modéliser un processus physique. C'est le cas d'un problème direct tel qu'il a été défini précédemment. Dans le cadre concret de la caractérisation par ellipsométrie, le problème est tout autre. En effet, le RN sera employé pour la résolution du problème inverse défini dans la partie I.3. C'est l'objet de la présente

partie où le RN sera dédié à la caractérisation optique et géométrique de couches minces de résine.

II.4.1 Position du problème

En rappelant le formalisme défini par l'équation I.71, le problème inverse consiste à inférer les paramètres \vec{a} à partir d'observations \vec{y} bruitées issues de mesures. L'objectif est ainsi de rechercher ou d'approcher la fonction F^{-1} .

De manière générale, si la relation qui lie les observations \vec{y} aux paramètres optogéométriques \vec{a} est univoque, le problème inverse est un problème de régression. La fonction $F^{-1}(\vec{y}, \vec{x})$ peut être ainsi approchée par une méthode de régression classique. Toutefois, la plupart sont linéaires (par exemple la méthode PLS [120]) et de ce fait peu performantes. L'utilisation des RN s'avère alors parfaitement adaptée. En effet, nous avons vu précédemment qu'ils constituent de véritables approximateurs non linéaires plus parcimonieux que bon nombre de modèles.

Le but du problème est d'associer à chaque observation \vec{y} l'espérance mathématique $E[\vec{a}/\vec{y}]$ par la fonction $F_{\vec{x}}^{-1}(\vec{y})$. Cette dernière peut être exprimée à l'aide d'une fonction $f_{W,Z}(\vec{y})$ relayée par un PMC à une couche cachée paramétrée par les poids W et Z :

$$s_m = f_{W,Z}(\vec{y}) = E[\vec{a}/\vec{y}] \quad (\text{II.22})$$

Le RN peut donc être envisagé pour donner une estimation de l'espérance mathématique de la variable aléatoire A pour un vecteur \vec{y} donné.

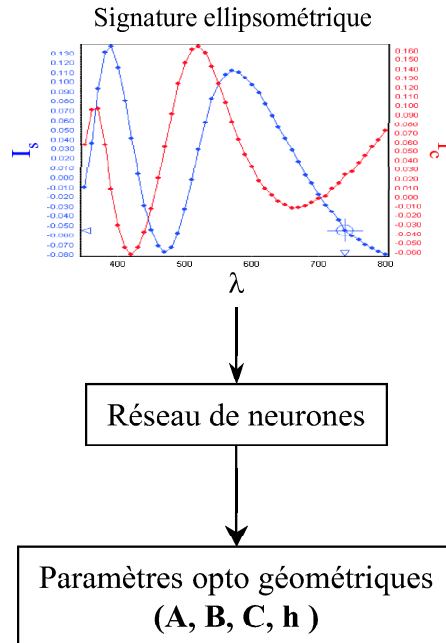


FIG. II.12: Représentation schématique de la caractérisation des propriétés optiques et géométriques de couches minces par réseau de neurones.

Par la suite, le PMC sera employé dans cet objectif dans le cadre précis de la caractérisation ellipsométrique pour la détermination des paramètres optiques et géométriques de couches minces de résine comme l'illustre schématiquement la figure II.12. Le but du

PMC est donc de trouver la relation qui existe entre la signature ellipsométrique mesurée et les paramètres de la couche, à savoir A , B , C (équation II.1) et h (figure II.1).

La signature ellipsométrique utilisée pour les études qui vont suivre sera constituée des intensités I_s et I_c mesurées dans la configuration II (équation I.36) de notre ellipsomètre. Le domaine spectral choisi est le visible ($400 \text{ nm} < \lambda < 800 \text{ nm}$ par pas de 20 nm). L'angle d'incidence est fixé à 60° (figure I.22). Le nombre d'entrées au réseau est donc celui du nombre de mesures (à savoir ici 42). Les sorties correspondent aux 4 paramètres à déterminer, à savoir les constantes de la loi de Cauchy A , B , C et l'épaisseur h de la couche.

II.4.2 Réseau de neurones dédié à la caractérisation ellipsométrique de couches minces de résine

La première étape consiste à effectuer la phase d'apprentissage du réseau. Cela nécessite la génération d'un corpus d'entraînement formé d'un certain nombre ($n_t = 2000$ par exemple) de couples $\{(I_s, I_c); (A, B, C, h)\}$. Les paramètres opto-géométriques sont compris dans des domaines bien définis. Ainsi, le PMC ne pourra caractériser efficacement que les échantillons dont les paramètres appartiennent à ces domaines.

Par exemple, définissons les domaines de variation suivants :

$$\begin{aligned} 1.57 &< A < 1.63 \\ 0.2 \text{ nm}^2 &< B < 1.2 \text{ nm}^2 \\ 0.4 \text{ nm}^4 &< C < 1.6 \text{ nm}^4 \\ 300 \text{ nm} &< h < 500 \text{ nm} \end{aligned} \tag{II.23}$$

Il est cependant plus significatif de voir dans quelles limites varie l'indice de réfraction. Pour cela, on représente sur la figure II.13 plusieurs lois de dispersion (environ 300) dont les paramètres sont tirés aléatoirement dans les domaines définis par l'équation II.23. En terme d'indice de réfraction, cela représente les différents matériaux que l'on peut caractériser avec le corpus d'entraînement défini.

Les données d'entraînement utilisées sont formées de 2000 échantillons simulés tirés aléatoirement de ces domaines. Ainsi, le corpus d'apprentissage contient $n_{entr} = 1400$ couples de l'effectif total, le corpus de validation en contient $n_{val} = 400$, et $n_{test} = 200$ pour le corpus test. Ce dernier est utilisé pour estimer les performances du réseau et n'intervient pas dans le processus d'apprentissage (voir partie II.3.3.2). Le nombre N_c de neurones cachés est de l'ordre de 15. Afin d'assurer une bonne généralisation, la méthode de l'arrêt prématuré (voir paragraphe II.3.3.3) est employée lors de la phase d'entraînement.

Les performances du PMC au cours de l'entraînement peuvent être visualisées sur la figure II.14, où l'on représente l'évolution des différentes erreurs quadratiques moyennes calculées sur les trois corpus présentés.

À la 328^{ème} itération, l'apprentissage s'arrête automatiquement lorsque E_{val} sur l'ensemble de validation commence à augmenter. Nous retrouvons la même allure de la partie gauche de la figure II.11 lorsque la technique d'arrêt prématuré est utilisée comme critère d'arrêt de l'algorithme d'apprentissage.

L'apprentissage du PMC avec des données purement théoriques sans bruit ne présente que peu d'intérêt pour caractériser des échantillons réels. Le cas idéal consisterait à en-

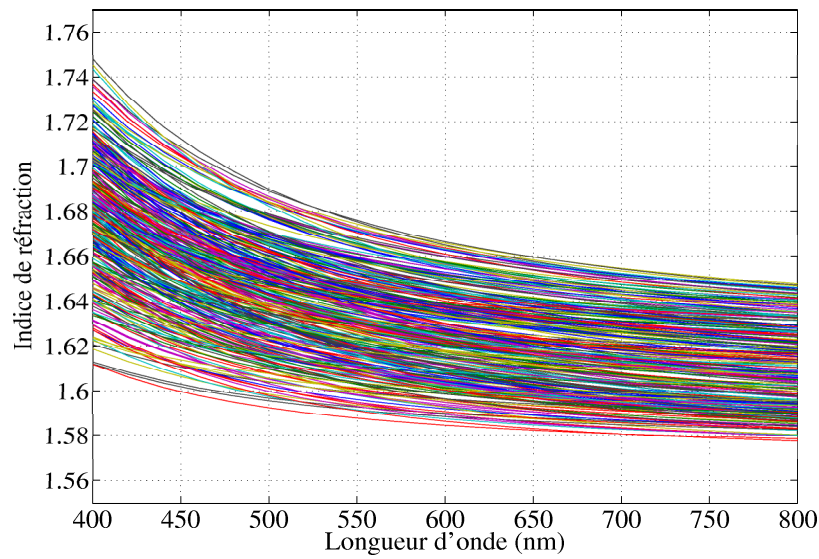


FIG. II.13: Représentation des lois de dispersion de 300 échantillons dont les paramètres sont compris dans les domaines de variation : $1,57 < A < 1,63$; $0,2 \text{ nm}^2 < B < 1,2 \text{ nm}^2$; $0,4 \text{ nm}^4 < C < 1,6 \text{ nm}^4$.

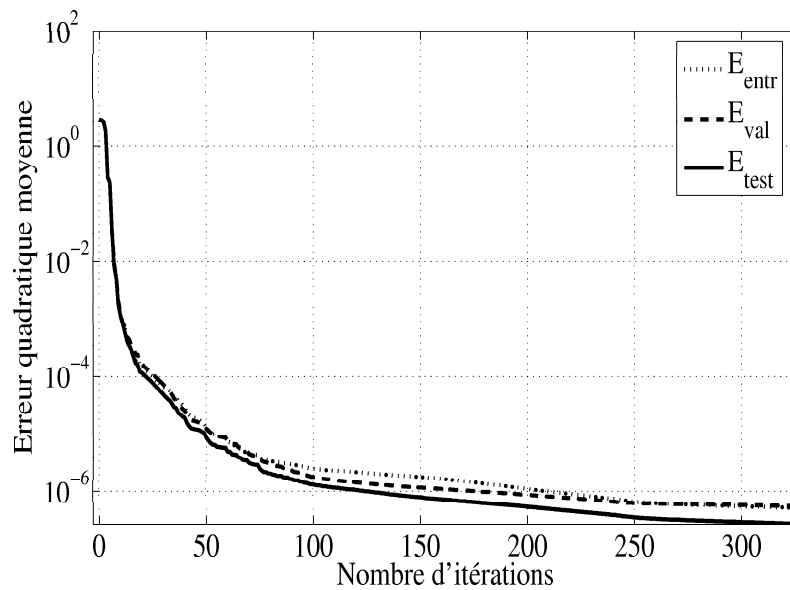


FIG. II.14: Evolution de l'erreur quadratique moyenne pondérée sur les paramètres au cours de l'apprentissage évaluée sur les corpus d'entraînement, de validation et de test jusqu'à l'arrêt de l'algorithme d'apprentissage par la méthode d'arrêt prématuré.

traîner le PMC avec des données issues de mesures expérimentales sur des structures de paramètres connus. Or, en pratique il est difficile voire presque impossible d'acquérir un grand nombre d'échantillons de l'ordre de celui utilisé pour l'entraînement. Ainsi, un bruit gaussien est ajouté à l'ensemble des points de mesures utilisés pour chaque échantillon simulé. Ce bruit supposé de distribution gaussienne possède une moyenne nulle et un écart type $\sigma = 3 \cdot 10^{-3}$ tel qu'il a été déterminé de façon simple dans le paragraphe I.3.4 du chapitre précédent. Le PMC est maintenant entraîné avec les nouvelles données bruitées.

Le tableau II.1 regroupe les erreurs quadratiques moyennes E_{test} obtenues pour les

paramètres A , B , C et h lorsque les données test sont bruitées ou théoriques. L'influence des données d'entraînement est également étudiée.

Données d'entraînement	Données test	Performance			
		E_{test}^A	E_{test}^B (nm ²)	E_{test}^C (nm ⁴)	E_{test}^h (nm)
non bruitées	non bruitées	$1,55 \cdot 10^{-5}$	$4 \cdot 10^{-4}$	$6 \cdot 10^{-4}$	0.1
non bruitées	bruitées	0,087	3,03	4,26	52,4
bruitées	bruitées	0,003	0,15	0,22	4

TAB. II.1: Performances du PMC calculées sur les données test lorsque celui-ci est entraîné sur des données purement théoriques ($\sigma = 0$), et sur des données bruitées ($\sigma = 3 \cdot 10^{-3}$).

Ainsi, un PMC entraîné sans bruit s'attend après entraînement à recevoir des nouvelles entrées non bruitées. Or, si ce n'est pas le cas le PMC ne fonctionne pas correctement. Comme le PMC sera dédié à la caractérisation des échantillons expérimentaux dont la mesure est naturellement bruitée, il est ainsi plus judicieux de l'entraîner avec des données bruitées.

Les différents paramètres A , B , C et h calculés par le PMC à partir des n_{test} signatures sont représentés sur la figure II.15, en fonction des paramètres exacts, dans le cas où le PMC est entraîné et testé avec des données bruitées. Une régression linéaire est effectuée afin d'évaluer la qualité du résultat.

Le PMC semble plus performant sur les paramètres A et h (figures II.15(a) et II.15(d)). En effet, ces deux paramètres sont plus influents sur la signature ellipsométrique que B et C . Ces deux derniers sont qualifiés de paramètres de second ordre, comme nous pouvons le constater à partir des figures II.15(b) et II.15(c).

Le paramètre d'intérêt étant l'indice de réfraction n , il est plus significatif de calculer l'erreur commise sur ce dernier. A partir des paramètres A , B , C estimés (figures II.15(a), II.15(b) et II.15(c)), nous avons calculés les lois de dispersions correspondantes (équation II.1) ainsi que l'erreur quadratique moyenne E_{test}^n entre les lois théoriques et estimées. La figure II.16 montre cette erreur pour l'ensemble des 200 échantillons test.

L'erreur sur l'indice est de l'ordre de $5 \cdot 10^{-3}$. Ce résultat ne doit pas être surprenant. En effet, il faut noter ici que l'erreur de mesure fixée pendant la simulation joue un rôle non négligeable sur la qualité du résultat. Une étude plus fine de cette erreur sera alors nécessaire pour améliorer significativement ces résultats. La mesure de cette dernière a volontairement été simplifiée et probablement surestimé précédemment. Toutefois, cette valeur d'erreur de $5 \cdot 10^{-3}$ sur l'indice est suffisante pour la plupart des applications en pratique.

D'après les résultats présentés jusqu'à présent, le PMC fonctionne correctement dans le domaine choisi et possède un bon niveau de généralisation.

II.4.3 Influence de la largeur des domaines d'entraînement

Les résultats satisfaisants de la partie précédente nous amènent à étudier plus précisément l'influence des domaines de variation des paramètres opto-géométriques.

La première étude concerne la variation du domaine de l'épaisseur h (D_i^h , $i=1\dots 8$ donnés dans le tableau II.2). Les autres domaines restent inchangés et identique au domaine initial (II.23).

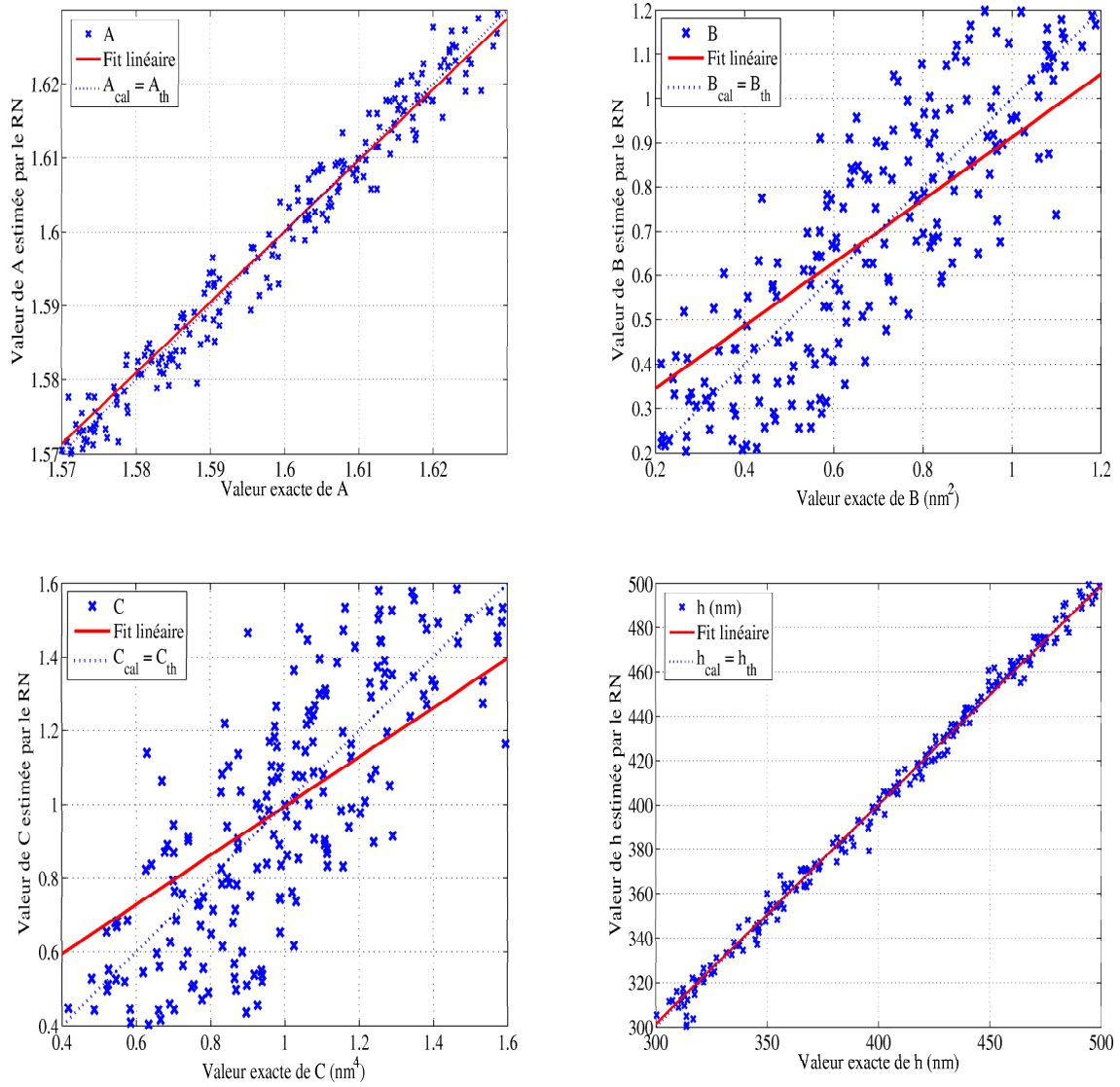


FIG. II.15: Représentation des différents paramètres opto-géométriques A , B , C et h calculés par le PMC à partir des signatures bruitées de 200 échantillons test en fonction de leurs paramètres exacts connus.

	D_1^h	D_2^h	D_3^h	D_4^h	D_5^h	D_6^h	D_7^h	D_8^h
Borne inférieur h_{inf} (nm)	340	315	290	265	240	215	190	165
Borne supérieur h_{sup} (nm)	390	415	440	465	490	515	540	565
Largeur du domaine (nm)	50	100	150	200	250	300	350	400

TAB. II.2: Définition des domaines de variation de l'épaisseur h .

Un PMC est entraîné sur chacun de ces domaines. L'architecture de ce dernier est gardée la même avec 15 neurones dans la couche cachée. La figure II.17 représente les erreurs normalisées des 4 paramètres (A , B , C , et h) données par la relation suivante :

$$E_N^i(D_i^h) = \frac{E_{test}^i(D_i^h)}{E_{test}^i(D_1^h)} \quad \text{pour } i = A, B, C \text{ et } h \quad (\text{II.24})$$

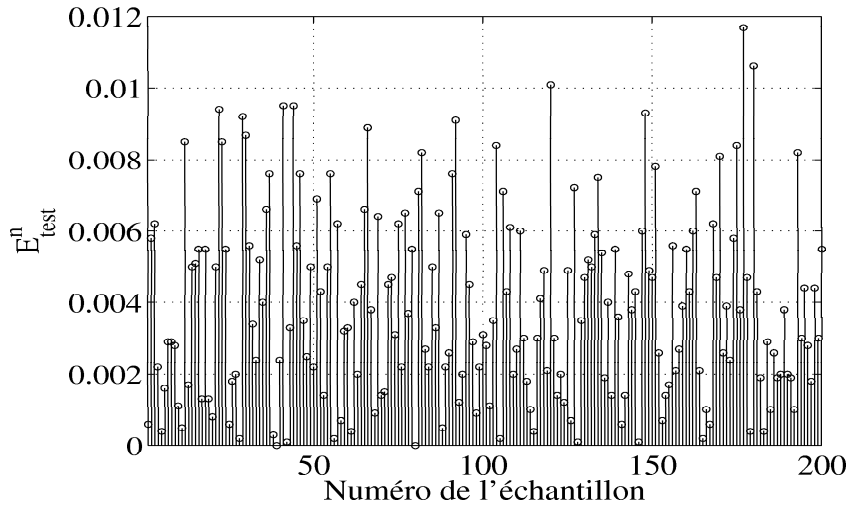


FIG. II.16: Erreur quadratique moyenne E_{test}^n sur l'ensemble des échantillons test.

sachant que les erreurs commises par le PMC sur le domaine D_1^h pour les 4 paramètres sont : $E_{test}^A = 0,0026$, $E_{test}^B = 0,15 \text{ nm}^2$, $E_{test}^C = 0,21 \text{ nm}^4$ et $E_{test}^h = 1,3 \text{ nm}$.

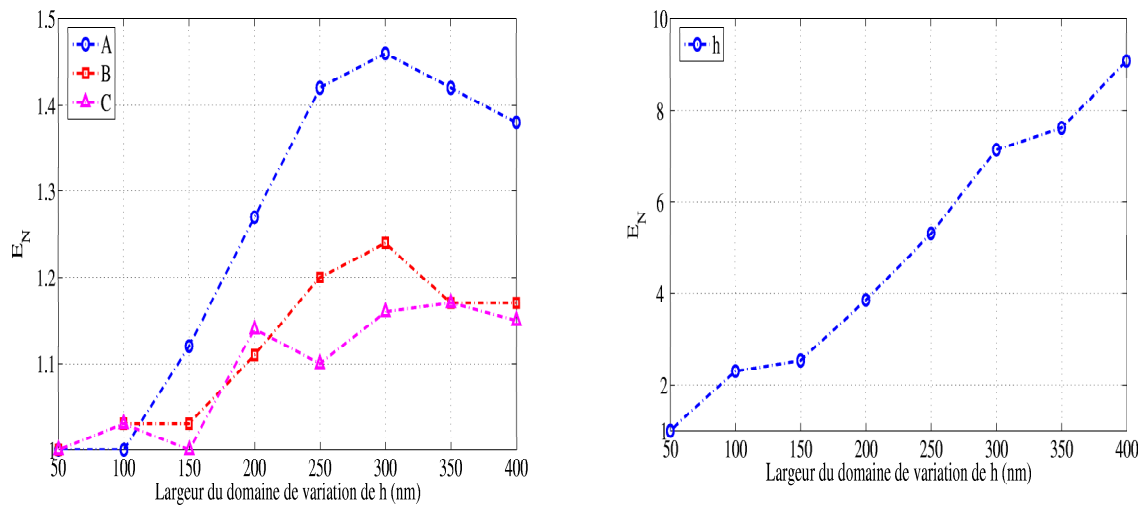


FIG. II.17: Variation relative des performances théoriques globales des paramètres A , B , C et h lorsque le PMC est entraîné pour différents domaines de h .

D'après les courbes de la figure II.17, l'élargissement du domaine de variation de h conduit logiquement à une dégradation des performances du RN. L'erreur sur h augmente fortement, tandis que les erreurs sur les 3 autres paramètres dont les domaines sont gardés fixes pendant l'entraînement subissent une légère augmentation.

La même opération est répétée en faisant varier cette fois-ci les domaines de tous les paramètres ensemble. Trois domaines sont choisis, notés D_1 , D_2 et D_3 , définis dans le tableau II.3.

Les réseaux de neurones, notés RN_1 , RN_2 et RN_3 , conservent toujours la même ar-

	A_{inf}	A_{sup}	B_{inf}	B_{sup}	C_{inf}	C_{sup}	h_{inf}	h_{sup}
D_1	1,59	1,61	0,4	1	0,6	1,4	300	500
D_2	1,57	1,63	0,1	1,3	0,3	1,7	200	600
D_3	1,55	1,65	0,1	1,9	0,1	2,2	100	700

TAB. II.3: Définition des trois domaines (D_1 , D_2 et D_3) de variation des paramètres A , B , C et h .

chitecture avec le même nombre de neurones. Les erreurs quadratiques calculés sur la base des n_{test} échantillons pour les 3 réseaux de neurones entraînés respectivement sur les domaines D_1 , D_2 et D_3 , sont regroupées dans le tableau II.4.

	E_{test}			
	E_{test}^A	E_{test}^B	E_{test}^C	E_{test}^h
RN_1	0,0021	0,12	0,15	4,4
RN_2	0,0041	0,2	0,26	16,7
RN_3	0,0069	0,32	0,46	32,7

TAB. II.4: Erreurs quadratiques moyennes E_{test} calculées sur chacun des paramètres (A , B , C et h) pour les 3 réseaux de neurones entraînés sur les domaines D_1 , D_2 et D_3 .

Contrairement au cas où le domaine d'un seul paramètre change, les performances globales du réseau se dégradent remarquablement lorsque tous les paramètres varient ensemble. Cela paraît évident puisque l'architecture du réseau reste la même face à la complexité croissante du problème.

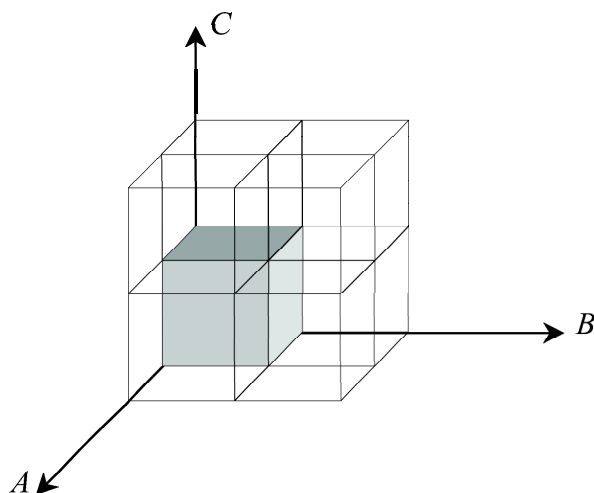


FIG. II.18: Représentation schématisée de l'espace engendré par 3 variables (A , B et C) dans un système d'axes orthonormés.

Pour illustrer graphiquement ce fait, considérons le cas où seulement trois paramètres sont à déterminer (A , B et C par exemple). L'espace engendré par les sorties de dimension 3 est représenté sur la figure II.18.

Si le domaine de chaque variable est doublé, l'espace total dans lequel le réseau sera amené à chercher la solution reste un cube mais qui possède un volume 8 fois plus grand

que le premier. Ainsi le problème devient beaucoup plus complexe. Pour garder des performances analogues, deux points doivent être vérifiés : d'une part, il faut assurer un bon nombre d'exemples d'entraînement suffisant pour que tout l'espace de recherche soit bien représenté, et d'autre part il faut aussi que le nombre de neurones soit suffisant pour modéliser efficacement la complexité du problème. Cela explique pourquoi les performances du réseau, ayant une architecture fixe, se dégradent lorsque les domaines deviennent de plus en plus grands.

Nous venons de montrer comment le RN pouvait être adapté à la résolution de problème inverse en ellipsométrie dans un cadre théorique. La définition préalable des domaines de variation des paramètres optogéométriques correspond au domaine de validité du RN et conditionne la qualité des résultats. En effet, les performances globales du RN dépendent de la dimension de l'espace représentatif des paramètres recherchés. Il est ainsi nécessaire d'assurer un bon compromis entre largeur du domaine et architecture du réseau (neurones cachés et nombre d'exemples d'entraînement). Il est à présent impératif d'étudier le degré de confiance associé à chaque sortie du RN. Ce point fera l'objet de la partie suivante.

II.5 Intervalle de confiance

Les erreurs quadratiques définies précédemment (paragraphe II.3.3.2) pour évaluer les performances du réseau de neurones sont calculées d'une manière globale et sont de ce fait indépendantes du vecteur d'entrée. Bien qu'elles permettent d'appréhender la qualité de l'apprentissage, cette information qualitative n'est pas suffisante dans un processus de caractérisation. Il faut impérativement estimer l'intervalle de confiance associé à chaque résultat. Nous avons vu au paragraphe I.3.4 le calcul de ces incertitudes lors d'une optimisation classique. Dans cette partie, la voie neuronale est privilégiée.

II.5.1 Estimation par réseau de neurones

La méthode que nous utiliserons pour estimer les intervalles de confiance consiste à réaliser un apprentissage de la variance des sorties s_m du PMC (II.12) sur les données d'entraînement [121].

Si on considère \vec{A} la variable aléatoire de densité de probabilité gaussienne et d'écart type σ_A et \vec{a} une réalisation particulière, le PMC permet de calculer $s = E[\vec{A}|\vec{y}]$. Trouver un intervalle de confiance associé à \vec{a} , revient à estimer σ_A . Ainsi, on pourra écrire que la valeur vraie est comprise entre $(s - \sigma_A)$ et $(s + \sigma_A)$ dans 68% des cas.

Considérons un processus que l'on doit modéliser à l'aide d'un PMC à partir de n_t couples d'observations (\vec{y}, \vec{a}) . On considère que \vec{y} et \vec{a} sont deux vecteurs de dimensions respectives N et N_s . Pour déterminer la variance s'_m sur la sortie s_m calculée par le modèle, on va procéder de la façon suivante (figure II.19) :

Un premier réseau de neurones (RN_1) de type PMC (figure II.6) est dédié à l'apprentissage de la relation liant \vec{y} à \vec{a} sur la base d'apprentissage comme explicité dans les paragraphes précédents. Le RN est optimisé, afin de donner de bonnes performances

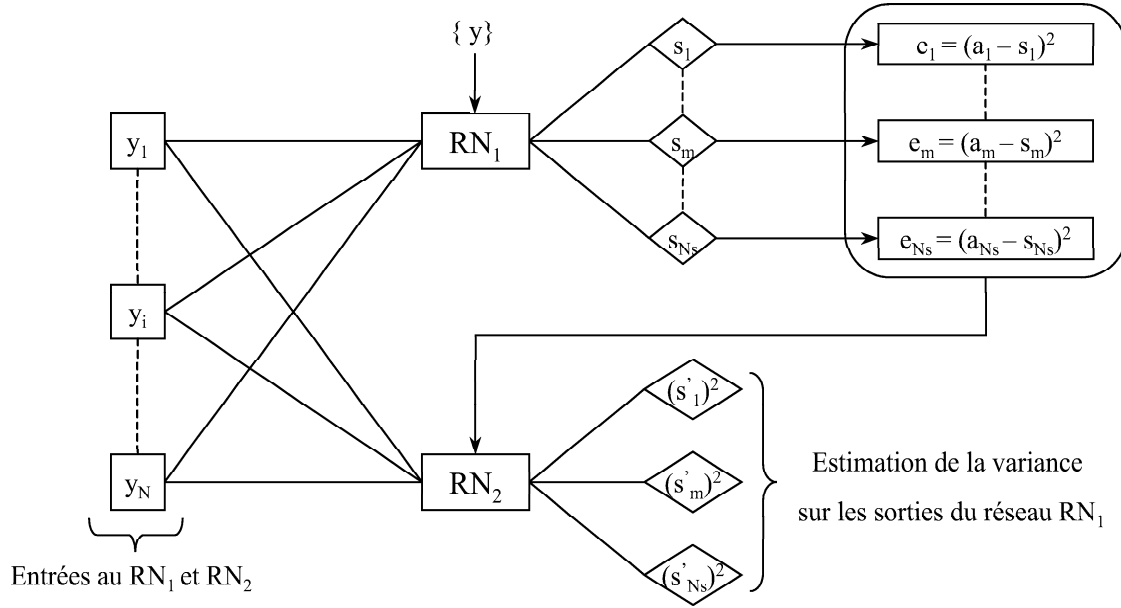


FIG. II.19: Représentation schématique du processus d'entraînement de la variance.

en généralisation à l'aide des méthodes énoncées auparavant. Nous avons vu que les sorties du réseau s_m représentent une bonne approximation de l'espérance mathématique de \vec{a} sachant \vec{y} , notée $E(\vec{A}/\vec{y})$. Une fois le PMC entraîné, il est facile de calculer l'erreur quadratique e_m réalisée sur chacune des sorties à l'aide des données d'entraînement par :

$$e_m = (a_m - s_m)^2 = (a_m - E[A_m/\vec{y}])^2 \quad (\text{II.25})$$

Ensuite, un deuxième réseau de neurones (RN_2) est mis en place pour inférer les erreurs e_m aux entrées \vec{y} . C'est un PMC de caractéristique similaire au précédent avec une fonction d'activation des neurones de sorties de type exponentielle afin de garantir des sorties positives. Les sorties du réseau RN_2 entraîné, notées $(s'_m)^2$, donnent une bonne approximation de l'espérance mathématique de e_m sachant \vec{y} , donc une estimation de la variance de a_m/\vec{y} :

$$s'_m = E[e_m/\vec{y}] = E[(a_m - E[A_m/\vec{y}])^2/\vec{y}] = \sigma_{A_m} \quad (\text{II.26})$$

Les réseaux RN_1 et RN_2 sont entraînés séparément, mais il est tout de même possible de réaliser cette procédure en utilisant un seul réseau de neurones spécifiquement dédié à cette tâche [121].

Dans le but d'illustrer la procédure d'apprentissage de la variance, nous allons donner un exemple de calcul simple. Considérons que l'on connaisse la formulation mathématique de $F_x^{-1}(\vec{y})$ permettant d'estimer \vec{a} supposé une variable aléatoire donnée par la forme suivante (\vec{a} et \vec{y} sont de dimension 1) :

$$E[a/y] = \sqrt{y} \sin(y) \quad (\text{II.27})$$

et ayant une distribution gaussienne :

$$\frac{1}{2\pi\sigma_A(y)} \exp \left[-\frac{(a - E[a/y])^2}{2\sigma_A^2(y)} \right] \quad (\text{II.28})$$

avec

$$\sigma_A(y) = 0,003 y^{2,5} \quad (\text{II.29})$$

En suivant la procédure décrite par la figure II.19, le RN_1 est employé pour apprendre la relation liant y et a . Ce dernier contient 3 neurones cachés et les données d'entraînement sont formées des couples (y, a) calculés pour tout y compris entre 2 et 15 avec un pas de 0,001. La dimension des vecteurs d'entrée et de sortie est égale à 1.

A l'aide de RN_2 contenant 10 neurones cachés, les erreurs commises par le premier réseau sont ajustées. Pour ce deuxième réseau, on emploie souvent un nombre de neurones plus grand puisque la fonction à modéliser est plus complexe que celle donnée au premier réseau.

La figure II.20 montre, pour un ensemble de points y contenu dans le domaine d'entraînement, les valeurs de a correspondantes calculées par le RN_1 , ainsi que les variances estimées par RN_2 en chacun de ces points.

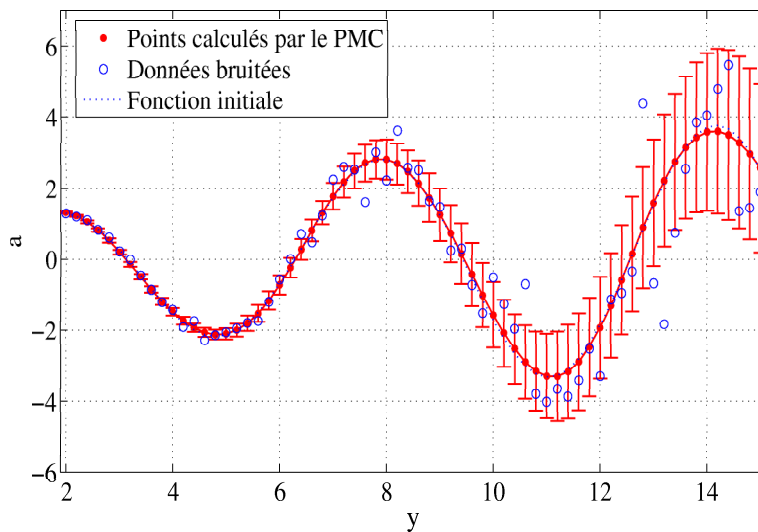


FIG. II.20: Résultats concernant l'apprentissage pour un RN de données bruitées (o) issues d'une fonction déterministe (en pointillé). Les sorties du RN_1 sont représentées (en traits pleins) agrémentées des sorties du RN_2 estimant les barres d'incertitude à 1σ .

La première chose à constater est que la fonction modélisée par RN_1 est bien ajustée à la fonction initiale non bruitée. Les intervalles de confiance associés aux points calculés par le réseau RN_2 sont donnés à 1σ . Ces derniers sont indépendants en chaque point et suivent l'allure de $\sigma_A(y)$. Ils donnent en plus une estimation ponctuelle de l'incertitude contrairement à l'erreur quadratique moyenne qui fournit une estimation globale qui a tendance à niveler les imprécisions locales du modèle. Ceci montre l'avantage du RN qui est un outil capable de donner un résultat ainsi que son degré de confiance.

II.5.2 Caractérisation de couches minces de résine

Tous les résultats précédents sont issus de simulations d'échantillons. Il est temps à présent de vérifier le comportement face à des données expérimentales. Considérons un échantillon simulé constitué d'une couche, dont les paramètres sont par exemple ($A = 1.585$, $B = 0.65 \text{ nm}^2$, $C = 0.85 \text{ nm}^4$, $h = 420 \text{ nm}$), déposée sur un substrat de verre BK_7 dont l'indice est connu (Annexe A). La signature ellipsométrique théorique correspondante (paragraphe II.4.1) est bruitée et introduite dans le RN entraîné au paragraphe II.4.2. Les paramètres estimés par le RN ainsi que leurs intervalles de confiance à 1σ sont affichés dans le tableau II.5.

Dans tous les cas, la barre d'incertitude englobe la valeur vraie de chaque variable aléatoire de sortie. On a également comparé la loi de dispersion théorique et celle obtenue à partir des paramètres calculés par le réseau, comme le montre la figure II.21.

	A	B(nm ²)	C(nm ⁴)	h(nm)
Valeur vraie	1,585	0,65	0,85	420
Valeur estimée par RN	1,586±0,001	0,7±0,15	0,9±0,13	416,6±4,7

TAB. II.5: Résultats de la caractérisation neuronale pour un échantillon simulé de paramètres optiques et géométriques (A , B , C et h) connus.

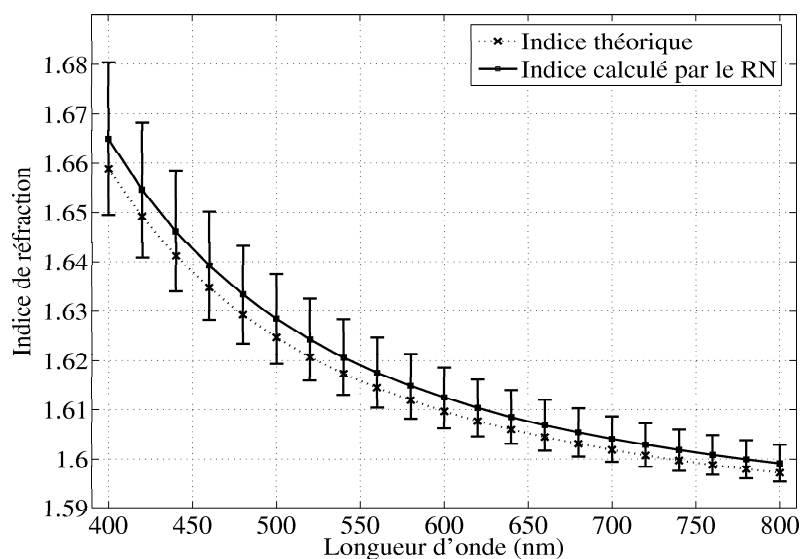


FIG. II.21: Loi de dispersion calculée par le RN pour l'échantillon simulé ($A=1.585$, $B=0.65 \text{ nm}^2$, $C=0.85 \text{ nm}^4$). Les barres d'incertitudes sont données à 1σ près.

Il n'est pas facile d'interpréter qualitativement l'erreur estimée sur les paramètres A , B et C . Par contre, il est plus parlant d'évaluer cette erreur sur l'indice optique N . Les lois, théorique et calculée, sont montrées sur la figure II.21 avec les incertitudes associées aux points donnés par le PMC déduites par la loi de propagation des incertitudes.

L'erreur moyenne trouvée est de l'ordre de $5 \cdot 10^{-3}$ entre la loi calculée par le réseau et la loi théorique. On remarque que l'incertitude sur n est plus grande aux basses longueurs d'onde. Il faut noter que les intervalles de confiance ainsi que l'erreur systématique observée peuvent être réduits par un apprentissage plus poussé du PMC voir une connaissance plus fine du bruit de mesure.

Après avoir vérifié que le réseau de neurones fonctionne correctement sur des réseaux simulés, nous allons à présent l'employer pour caractériser une couche mince réelle. Les résultats seront comparés avec ceux donnés par une autre méthode d'optimisation classique, à savoir la méthode de LM.

L'échantillon expérimental est conforme au modèle de la figure II.1, et est constitué d'une couche de résine déposée sur un substrat de verre d'indice connu (annexe A). Le détail de la signature ellipsométrique mesurée expérimentalement ainsi que le RN utilisé sont spécifiés dans la partie II.4.1. Les résultats donnés par le PMC sont reportés dans le tableau II.6 et ceux relatifs à l'optimisation par LM pour différents points de départ sont regroupés dans le tableau II.7.

	A	B(nm ²)	C(nm ⁴)	h(nm)
RN	1,609±0,003	0,68±0,14	0,98±0,17	363,4±1,5

TAB. II.6: Résultats obtenus par la caractérisation neuronale (RN).

Essai	Point de départ h_i (nm)	Résultats LM			
		A	B	C	h (nm)
1	300	1,67 ±0,096	-1,03±6,1	-4,88±8,23	362±11,1
2	350	1,609±0,003	0,71±0,22	0,81±0,3	364,8±0,3
3	400	1,609±0,003	0,68±0,22	0,84±0,3	364,8±0,3
4	450	1,624±0,004	-0,32±0,3	2,2±0,38	365±0,4
5	500	1,742±0,049	-3,7±3,35	2±4,9	566,7±7,1

TAB. II.7: Résultats obtenus par optimisation classique (méthode de Levenberg-Marquardt) pour plusieurs épaisseurs initiales h_i ; les valeurs initiales de A, B et C sont fixes et respectivement égales 1.6, 1 et 1 proche de la solution finale.

Il est évident que dans le cas de l'optimisation classique le résultat dépend du point de départ choisi. Les essais 2, 3 et 4 convergent vers la solution optimale qui concorde avec celle donnée par le PMC. Pour les essais 1 et 5, on peut attribuer la mauvaise convergence à la présence de minima locaux. Quant au PMC, il trouve la solution optimale et résiste mieux à ces minima présents dans son domaine d'entraînement.

II.6 Choix de la signature optique optimale

Le choix des entrées d'un RN influe directement sur la qualité du résultat donné par celui-ci. En effet, la complexité de la surface d'erreur (représentée par χ^2) est directement liée aux observations prises comme entrées. Lors de la modélisation d'un processus donné, les entrées utilisées doivent donc contenir les informations nécessaires et utiles. Ainsi, l'ajout d'entrées considérées comme non pertinentes pour le modèle, voire indépendantes des sorties à modéliser, n'apporte aucune information supplémentaire et peut même, le cas échéant, dégrader la qualité du résultat. Par définition [122], une variable pertinente est une variable telle que sa suppression entraîne une détérioration des performances (qualité de prédiction en régression) du modèle. Dans le même ordre d'idée, les petits réseaux possédant moins de poids ont tendance à être plus performant et généralisent mieux que les grands réseaux [123].

Remarque : Nous devons préciser que le problème de choix des variables descriptives n'est pas unique au RN, mais à toute modélisation. Il existe dans la littérature de nombreuses méthodes statistiques que nous ne détaillerons pas ici. Nous nous concentrerons par la suite sur des méthodes propres aux RN.

Cette partie est consacrée au problème de la "sélection de modèles". L'idée consiste d'une part à réduire le nombre de variables en supprimant les non pertinentes, et d'autre part à choisir un nombre de neurones optimal aboutissant à une architecture neuronale la plus simple garantissant des performances souhaitées. Il faut noter que le premier point est différent de celui de la réduction de l'espace des variables décrit au paragraphe II.3.2.3) qui consiste à définir de nouvelles variables à partir de celles choisies au départ.

Le début de cette partie expose quelques méthodes qualifiées d'élagage des entrées. Elles permettent d'analyser les performances du réseau lorsqu'une connexion est supprimée, et de rechercher ainsi simultanément l'architecture optimale et le nombre d'entrées pertinentes. Nous nous intéressons ensuite plus particulièrement à une méthode de sélection appelée HVS (en anglais *Heuristic Variable Selection*) qui sera utilisé dans la suite de ce travail, et appliquée à la caractérisation de couches minces de résine.

II.6.1 Méthodes classiques d'élagage

La méthode d'élagage (de l'anglais *pruning*) la plus connue est la méthode appelée OBD (en anglais *Optimal Brain Damage*). Elle est introduite par LeCun [124] comme une méthode pour réduire significativement le nombre des poids dans un réseau de neurones. En général, partant d'un modèle complet, on estime la variation de la fonction coût lorsqu'un poids particulier est supprimé du réseau. Celui qui produit la plus faible variation est supprimé.

La procédure d'élagage commence une fois que l'apprentissage est terminé. La fonction coût est supposée quadratique, pour que sa variation δC puisse être approchée par un développement en série de Taylor :

$$C(\xi + \delta\xi) - C(\xi) = \delta C = \left(\frac{\partial C}{\partial \xi} \right)^T \delta\xi + \frac{1}{2} \delta\xi^T H \delta\xi \quad (\text{II.30})$$

avec ξ un vecteur regroupant tous les paramètres ajustables du RN, à savoir tous les éléments contenus dans W et Z , et $H = \frac{\partial^2 C}{\partial \xi^2}$ est la matrice Hessienne correspondante.

Après apprentissage, la fonction coût est minimale (les dérivées s'annulent), on a ainsi :

$$\delta C = \frac{1}{2} \delta\xi^T H \delta\xi \quad (\text{II.31})$$

et donc si on suppose la matrice H diagonale :

$$OBD(\xi_i) = \delta C(\xi_i) = \frac{1}{2} \xi_i^2 \frac{\partial^2 C}{\partial \xi_i^2} \quad (\text{II.32})$$

Cette quantité est également appelée la sensibilité du poids ξ_i . A l'aide de la formule II.32, les sensibilités de tous les poids sont calculées en supprimant les poids ayant les sensibilités les plus faibles. Cette méthode permet aussi la suppression des entrées qui ne se trouvent reliées à aucun neurone. Elle permet ainsi de rechercher l'architecture optimale

du réseau et les entrées pertinentes du modèle. Un réapprentissage est ensuite effectué afin de compenser l'effet de la suppression.

Une amélioration de la méthode OBD est la méthode appelée OBS [125] (en anglais *Optimal Brain Surgeon*) qui ne fait pas l'hypothèse que la matrice Hessienne soit diagonale. La sensibilité d'un poids est ainsi calculée par :

$$OBS(\xi_i) = \frac{1}{2} \frac{\xi_i^2}{(H^{-1})_{ii}} \quad (\text{II.33})$$

Dans cette technique, le poids ayant la plus petite sensibilité est supprimé et les autres poids sont ajustés à l'aide de la formule suivante :

$$\xi'_i = \xi_i - \frac{\xi_i}{(H^{-1})_{ii}} H^{-1} v_i \quad (\text{II.34})$$

avec v_i est un vecteur dont la seule composante i est à 0 et les autres à 1.

D'autres auteurs [126] ont proposé des extensions d'OBD et d'OBS, à savoir EBD (*Early Brain Damage*) et EBS (*Early Brain Surgeon*), qui peuvent être utilisées avec l'arrêt prématuré.

II.6.2 Méthode de sélection HVS

La méthode choisie dans notre travail, à savoir celle basée sur la mesure HVS (en anglais Heuristic Variable Selection), sera présentée ici en détail. Cette méthode, qualifiée d'ordre 0, peut être appliquée non seulement pour la sélection de variables mais aussi pour la recherche d'une architecture optimale du PMC. Elle a été développée au laboratoire LOCEAN de l'université Paris VI [127, 128]. Pour les lecteurs intéressés, la référence [129] fournit en détail un ensemble de méthodes utilisées dans ce cadre.

Considérons le PMC de la figure II.6, avec N entrées, une couche cachée de N_c neurones et N_s sorties. La valeur de la connexion entre l'entrée i ($i = 1, 2, \dots, N$) et le neurone j , notée $w_{j,i}$, reflète l'importance de leur lien. Comme cette valeur peut être positive ou négative selon que la connexion soit excitatrice ou inhibitrice, la force de cette connexion est quantifiée par $|w_{j,i}|$. On définit, dans ce cas, la *contribution partielle* $\pi_{j,i}^e$ de chaque entrée i sur le neurone caché j comme une proportion de toutes les autres forces de connexions arrivant à j .

$$\pi_{j,i}^e = \frac{|w_{j,i}|}{\sum_{k=1}^N |w_{j,k}|} \quad (\text{II.35})$$

Le but est d'estimer la contribution totale de l'entrée i sur les sorties par les contributions partielles des neurones de la couche cachée. Comme chaque neurone ne contribue pas de la même façon, nous définissons ainsi la *contribution relative* S_i de l'entrée i sur les sorties comme une somme de ses contributions partielles pondérées par l'importance des neurones cachés.

$$S_i = \sum_{j=1}^{N_c} \pi_{j,i} \delta_j \quad (\text{II.36})$$

δ_j étant la contribution relative du neurone caché j sur les N_s sorties calculée par le

même formalisme en imposant une importance de 1 aux neurones de sorties :

$$\delta_j = \sum_{m=1}^{N_s} \pi_{m,j}^s \delta_m \quad \text{avec} \quad \pi_{m,j}^s = \frac{|z_{m,j}|}{\sum_{k=1}^{N_c} |z_{m,k}|} \quad \text{avec} \quad \delta_m = 1 \quad (\text{II.37})$$

Cette méthode est de type *backward selection*, où les variables sont supprimées une à une jusqu'à la dernière variable. La procédure de sélection se déroule de la façon suivante : Etant donné un problème à N variables descriptifs, un premier apprentissage est effectué en employant l'arrêt prématuré. La pertinence de chaque variable est calculée selon l'équation II.36. L'entrée la moins influente sur les sorties du réseau, estimée par la contribution relative S_i la plus faible, est alors supprimée. L'apprentissage est ensuite poursuivi, en prenant comme point de départ des différents poids la solution trouvée par l'entraînement précédent. Cette démarche est répétée jusqu'à ce que toutes les entrées soient supprimées. On se retrouve finalement avec N modèles différents (représentés par N PMC). Le réseau optimal correspond à celui qui possède l'erreur minimale en fonction du nombre de variables.

Cette méthode est facile à mettre en œuvre et s'applique à un PMC ayant un nombre quelconque de couches. Elle sera également utilisée pour ajuster l'architecture du réseau (optimisant le nombre N_c de neurones cachés) en l'appliquant aux neurones de la couche cachée [128]. Le principe est le même en considérant simplement chaque sortie des neurones cachés comme une entrée d'un sous-réseau composé de la couche de sortie et de toutes les couches cachées intermédiaires. Ainsi, l'optimisation de l'architecture du réseau devient un problème identique à celui de sélection de variables, dans le sous-réseau en question.

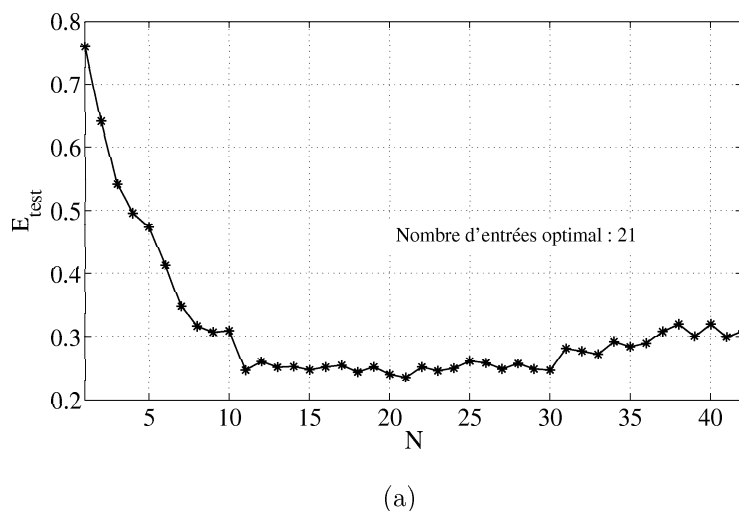
II.6.3 Signature ellipsométrique optimale dans le cas de couches minces de résine

La méthode HVS peut ainsi être utilisée pour la sélection d'une signature ellipsométrique optimale utilisée pour la caractérisation de couches minces de résine. Elle sera également employée pour réduire la dimension du PMC conservant un nombre de neurones cachés optimal.

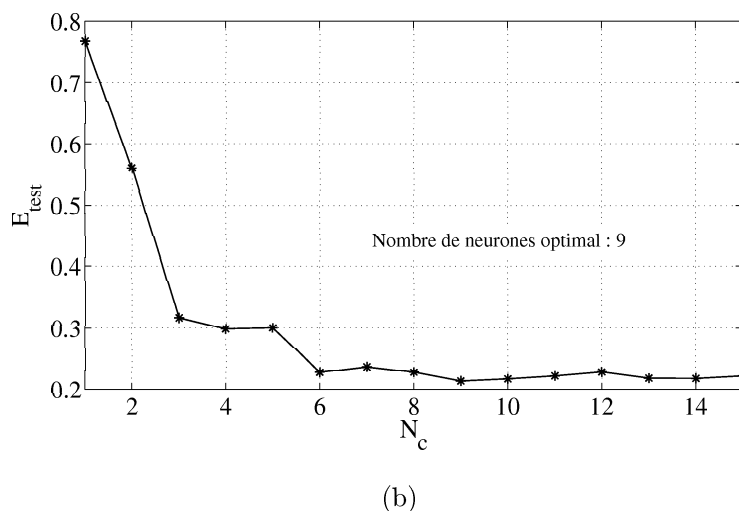
Reprenons les mêmes conditions définies dans la partie II.4.2 pour la caractérisation de couches de résine. Le RN initial est composé de 42 entrées, 15 neurones cachés et 4 sorties (A , B , C et h). La méthode HVS est appliquée en suivant la même procédure de sélection décrite dans le paragraphe précédant, pour les entrées et pour les neurones cachés.

L'évolution de l'erreur quadratique moyenne E_{test} , calculée à chaque fois qu'une entrée est supprimée du modèle initial, est représentée sur la figure II.22(a). Par soucis de clarté, l'axe des abscisses est tracé à l'envers. En effet, le dernier point (d'abscisse 42) correspond au cas où le RN contient toutes les entrées descriptives. Ces dernières sont ensuite supprimées une à une jusqu'à la dernière, correspondant au point d'abscisse 1 de la courbe.

Lorsque les entrées sont nombreuses, l'erreur subit une légère augmentation. Ceci est dû tout simplement à la complexité de l'architecture du RN lorsque le nombre de paramètres est très grand, à savoir ici $42 \times 15 \times 3 = 1890$ paramètres ajustables. On peut noter que la suppression des 32 premières entrées n'affecte pas l'erreur quadratique moyenne et au contraire l'améliore sensiblement. Théoriquement, nous devons chercher le nombre



(a)



(b)

FIG. II.22: Evolution de l'erreur quadratique moyenne en fonction du nombre des (a) entrées N du nombre de (b) neurones cachés N_c .

optimal correspondant au vrai minimum de la courbe d'erreur. Or, on distingue clairement sur la figure une zone pour laquelle l'erreur E_{test} semble quasiment constante et optimale (entre 10 et 30). Le choix du nombre de variables descriptives n'est donc en pratique pas critique. Ainsi, en tenant compte du compromis performance et rapidité de mesure nous désignerons le nombre de variable optimal par la valeur basse de cette zone.

Pour fixer les idées nous avons détaillé dans le tableau dans le tableau II.8 les mesures composant la signature optique sélectionnée estimées selon leur ordre de pertinence.

A partir de ces intensités sélectionnées formant la nouvelle signature ellipsométrique comme entrée au réseau, on procède à la sélection de l'architecture optimale en considérant le sous RN composé d'une seule couche de 4 neurones et dont les entrées sont les sorties des neurones cachés. Les résultats sont représentés sur la figure II.22(b).

De la même façon que précédemment, on choisit le nombre de neurones correspondant à une erreur E_{test} optimale. Ainsi, afin d'assurer un bon compromis entre une complexité de RN raisonnable et une erreur faible, le nombre de neurones choisie est $N_c = 6$.

La nouvelle architecture du PMC qualifié d'optimal par la méthode HVS, est composée

$\lambda(nm)$	400	420	440	460	480	500	520	540	560	580	600
I_s					● ₈	● ₁₀					
I_c	● ₁₁	● ₇					● ₁			● ₄	
$\lambda(nm)$	620	640	660	680	700	720	740	760	780	800	
I_s				● ₉			● ₆				
I_c	● ₃			● ₅				● ₂			

TAB. II.8: Liste des intensités (I_s , I_c) sélectionnées par la méthode HVS. L'importance estimée de chacune d'elle est donnée en indice (●₁ par exemple correspond à l'intensité la plus influente.)

de 11 entrées, 6 neurones et 4 sorties. Un nouvel entraînement est réalisé avec cette fois-ci 90% de l'effectif total (pas de corpus de validation), le corpus test étant composé des 10% restants. Rappelons que la méthode de l'arrêt prématuré (paragraphe II.3.3.3) permet d'arrêter la procédure d'apprentissage pour un PMC d'architecture surdimensionnée afin d'éviter un surapprentissage éventuel néfaste en terme de généralisation. Or, dans ce cas, l'architecture sélectionnée par la méthode HVS étant supposée de taille optimale pour une bonne généralisation, la méthode de l'arrêt prématuré ne s'impose plus.

La couche de résine est maintenant caractérisée avec le nouveau PMC simplifié. Les paramètres optiques et géométriques de la couche (A , B , C et h) estimés par ce PMC sont donnés dans le tableau II.9. Ces derniers concordent bien avec ceux obtenus précédemment par le réseau complet (tableau II.6).

	A	B(nm ²)	C(nm ⁴)	h(nm)
RN optimal	1,608±0.003	0,62±0,18	1,02±0,22	365,4±2,4

TAB. II.9: Caractérisation des paramètres A , B , C et h de la couche de résine par le PMC optimal sélectionné par la méthode HVS (11 entrées et 6 neurones).

L'utilisation de la méthode HVS nous permet de réduire considérablement le nombre d'entrées, et par la même occasion le temps nécessaire pour la mesure de la signature optique. D'un autre côté, l'avantage du choix d'une architecture optimal est de diminuer le temps d'entraînement du réseau ainsi que le risque de convergence prématurée puisque le nombre de paramètres à optimiser est beaucoup plus petit. Le temps global de caractérisation se trouve donc grandement réduit. Toutefois, il faut noter que les résultats de la méthode HVS dépendent des conditions du problème comme par exemple le choix du domaine d'entraînement. La validité des résultats fournis est donc limitée aux conditions expérimentales choisies. De plus, l'initialisation des poids étant aléatoire, la valeur de ces derniers après la première itération d'une sélection HVS est différente à chaque essai. Ainsi, il se peut que la composition de la signature optique optimale s'en trouve modifiée. En effet, il est fort probable que certaines intensités possèdent des contributions analogues sur les sorties et la sélection d'une seule fournit les informations suffisantes. Or, ce n'est pas systématiquement la même qui peut être sélectionnée.

Nous verrons dans le chapitre 3 l'utilité de cette méthode pour la scatterométrie où le calcul direct est beaucoup plus long que celui utilisé pour les couches minces.

Dans ce que nous venons de voir jusqu'à présent, le RN était appréhendé comme un régresseur non linéaire focalisé plus particulièrement sur la résolution d'un problème inverse. Un PMC est donc capable, à partir de la mesure d'une signature ellipsométrique fournie

en entrée, d'estimer en sortie les paramètres (optiques ou géométriques) caractérisant la structure sous test. Il peut également estimer l'intervalle de confiance (partie II.5.1) lié à chacun de ces paramètres. La partie suivante présente une autre application différente des réseaux de neurones, la classification et les applications possibles en métrologie optique.

II.7 Classification par réseaux de neurones

L'objectif de cette partie est de montrer une autre application des réseaux de neurones dans le domaine de la caractérisation ellipsométrique, à savoir la classification. Après avoir présenté son intérêt en pratique, nous allons détailler le formalisme mathématique classique. Puis, nous exposerons la classification par voie neuronale pour finalement montrer son apport dans le domaine de la caractérisation des couches minces.

II.7.1 Intérêt de la classification dans la caractérisation optique

Le réseau de neurones est également sollicité pour résoudre des problèmes de *classification*. Etant donné un ensemble d'observations, la classification consiste à attribuer à chacune d'elle une classe parmi plusieurs.

Dans notre cas, lors de la résolution du problème inverse, la première étape consiste à fixer un certain nombre d'hypothèses (géométriques ou autres) défini par un certain nombre de paramètres qui décrivent au mieux le système étudié (symbolisé par le vecteur \vec{x} d'après les notations utilisées auparavant). Dans le cas de l'ellipsométrie, il s'agit principalement de l'hypothèse structurale de l'échantillon. Ainsi, pour la caractérisation des couches minces, on suppose d'abord un modèle géométrique composé d'une couche d'épaisseur h sur un substrat d'indice connu. En scatterométrie, l'hypothèse classique de départ consiste à fixer le profil géométrique le plus probable. Dans le cas d'une caractérisation par RN qui nous intéresse ici, il peut également s'agir du domaine de validité des paramètres d'intérêt qui définissent les limites du modèle neuronal.

L'hypothèse structurale du modèle structural est l'essence même des méthodes d'optimisation, que le problème soit simple ou complexe. En effet, les paramètres \vec{a} recherchés sont définis suivant les hypothèses choisies. Il se peut parfois que la structure supposée soit fautive et ne corresponde pas physiquement à celle étudiée. On dit dans ce cas que le problème est mal posé, ce qui aboutira à des résultats erronés. Ceci reste le point faible de la plupart des méthodes d'optimisation qui restent assez peu flexibles de ce point de vue.

Dans le problème qui nous concerne, l'idée est d'utiliser la signature optique pour obtenir une première information qualitative sur les hypothèses à choisir lors de la phase de résolution du problème inverse. La figure II.23 illustre dans le cas général la classification d'une signature ellipsométrique dans une classe parmi K . Chaque classe peut représenter un modèle avec différentes hypothèses de départ (à savoir différents vecteurs \vec{x}).

Il est ainsi possible de concevoir, avant la phase de caractérisation, un classifieur permettant de choisir l'hypothèse structurale (présence d'une rugosité de surface ou non, présence d'une couche en plus ou non, etc. . .). Dans la caractérisation des couches minces par exemple, il est nécessaire d'avoir une idée de l'épaisseur de la couche pour bien choisir le point de départ pour la plupart des méthodes d'optimisation classique au risque de converger dans un minimum local. Ainsi, à titre d'exemple nous considérerons l'applica-

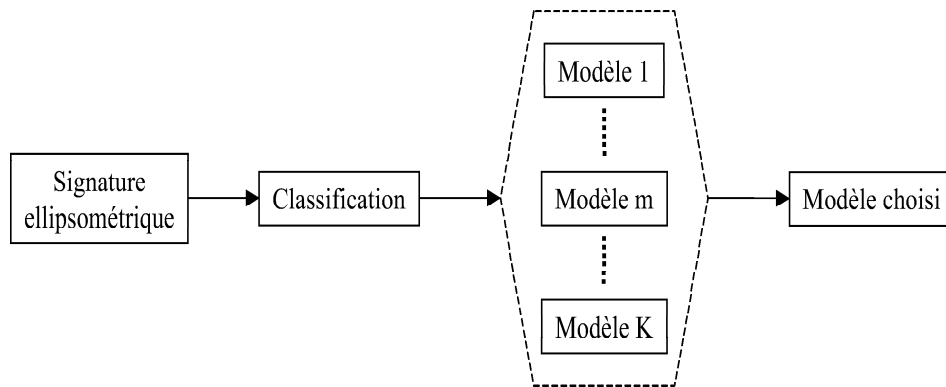


FIG. II.23: Schéma illustrant le phénomène de classification d'une signature ellipsométrique selon K classes distinctes.

tion qui consiste à localiser une gamme d'épaisseur à partir de la signature ellipsométrique d'un échantillon.

II.7.2 Le problème de la classification probabiliste

Dans le formalisme probabiliste de la classification, les descripteurs \vec{y} et les classes C_k ($k = 1, \dots, K$) sont définis par des variables aléatoires respectivement à valeurs continues et discrètes. On définit ainsi la densité de probabilité conditionnelle de chaque descripteur \vec{y} de la classe k par : $f(\vec{y}|k)$. On cherche donc la probabilité que la variable aléatoire définissant la classe soit égale à un numéro de classe (k par exemple). C'est la probabilité a posteriori, notée $P_r(C_k|\vec{y})$. On définit également la probabilité a priori de la classe C_k , notée $P_r(C_k)$, comme la probabilité de tirer au hasard un échantillon de la classe k parmi la population d'objets à classer.

Pour résoudre un problème de classification on distingue en général deux catégories de méthodes :

- Les méthodes indirectes, qui permettent de calculer les probabilités *a posteriori* à condition que les probabilités *a priori* et les densités de probabilité conditionnelles de chaque classe soient préalablement connues. Le théorème de Bayes est l'une des méthodes basiques en statistique appartenant à cette famille.
- Les méthodes directes, estiment directement les probabilités *a posteriori* d'appartenance aux classes en évitant de passer par la phase d'estimation des densités de probabilité. Les réseaux de neurones, dont le principe de fonctionnement en tant que classifieurs sera détaillé par la suite, font partie de ces méthodes.

Pour les méthodes indirectes, une estimation précise des densités de probabilité de classes est difficile, voire hasardeuse. Cependant, il existe deux types de méthodes permettant d'estimer les lois de probabilité : les méthodes paramétriques et les méthodes non paramétriques. Les premières supposent que les lois de probabilités obéissent à une loi de type gaussienne par exemple. Les secondes permettent de converger vers les distributions des classes sans aucune hypothèse, une des plus anciennes est la méthode de Fenêtres de Parzen [130]. Pour plus de détails sur ce sujet, le lecteur intéressé peut consulter la référence [131].

Théorème de Bayes Ce théorème est un résultat essentiel en théorie des probabilités. Considérons le problème où l'on doit classer un ensemble d'objets, appartenant à une population donnée, en K classes distinctes. Le théorème de Bayes permet d'exprimer la probabilité *a posteriori* $P_r(C_k|\vec{y})$ qu'un individu y appartienne à la classe k , en considérant la probabilité *a priori* des individus des différentes classes et de leur distribution dans l'espace des descripteurs, à l'aide de la formule suivante :

$$P_r(C_k|\vec{y}) = \frac{P_r(C_k) \cdot f(\vec{y}|C_k)}{\sum_{j=1}^K P_r(C_j) \cdot f(\vec{y}|C_j)} \quad (\text{II.38})$$

Ainsi, la règle décision de Bayes consiste à affecter à l'objet la classe dont la probabilité *a posteriori* est la plus grande. Il a été démontré que la décision donnée par ce théorème minimise la probabilité ou le risque d'erreur de classement [132], sachant que cette dernière n'est pas nulle.

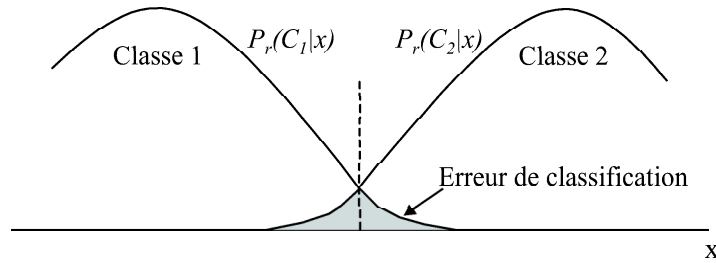


FIG. II.24: Interprétation géométrique de la règle de décision de Bayes ; la partie grise représente la probabilité de commettre un mauvais classement.

Ce théorème est illustré géométriquement par la figure II.24, où l'on représente les densités de probabilités correspondantes à deux classes C_1 et C_2 , notées respectivement $P_r(C_1|\vec{y})$ et $P_r(C_2|\vec{y})$. Le trait vertical correspond au seuil de classification donné par la règle de décision de Bayes entre les deux classes. En prenant cette décision, l'erreur d'affecter l'objet à la classe 1 alors qu'il est de classe 2 n'est pas nulle mais minimisée (ou vice-versa).

Nous allons à présent illustrer ce théorème à l'aide d'un exemple physique simple à une dimension. Pour cela, considérons une série d'échantillons (par exemple des couches minces de résine déposées sur un substrat) issue d'une chaîne de fabrication, destinée à produire des couches ayant deux épaisseurs différentes, $h_1 = 200$ nm et $h_2 = 250$ nm. On suppose que le paramètre épaisseur pour chaque groupe est vu comme une variable aléatoire est de distribution gaussienne d'écart-type $\sigma_{h_1} = 20$ nm et $\sigma_{h_2} = 25$ nm.

On souhaite, à la sortie de la chaîne, et à partir d'une simple mesure d'épaisseur de couche séparer les échantillons en deux catégories selon l'épaisseur obtenue. Les classes qui doivent contenir les couches d'épaisseur voisine de h_1 et de h_2 sont notées respectivement C_1 et C_2 . Supposons que la distribution des épaisseurs de la première catégorie d'échantillons suive une loi gaussienne de moyenne 200 nm et d'écart-type 20 nm ; alors la densité de probabilité de la variable h est donnée par :

$$f(h|C_1) = \frac{1}{\sigma_{h_1} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{h - h_1}{\sigma_{h_1}} \right)^2 \right] \quad (\text{II.39})$$

et que celle de la deuxième catégorie soit aussi une gaussienne de moyenne 250 nm et d'écart-type 25 nm :

$$f(h|C_2) = \frac{1}{\sigma_{h_2}\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{h - h_2}{\sigma_{h_2}} \right)^2 \right] \quad (\text{II.40})$$

Les densités de probabilité sont représentées sur la figure II.25. Il est clair que la seule variable descriptive choisie, qui est l'épaisseur, n'est pas totalement discriminante car les deux gaussiennes se recouvrent.

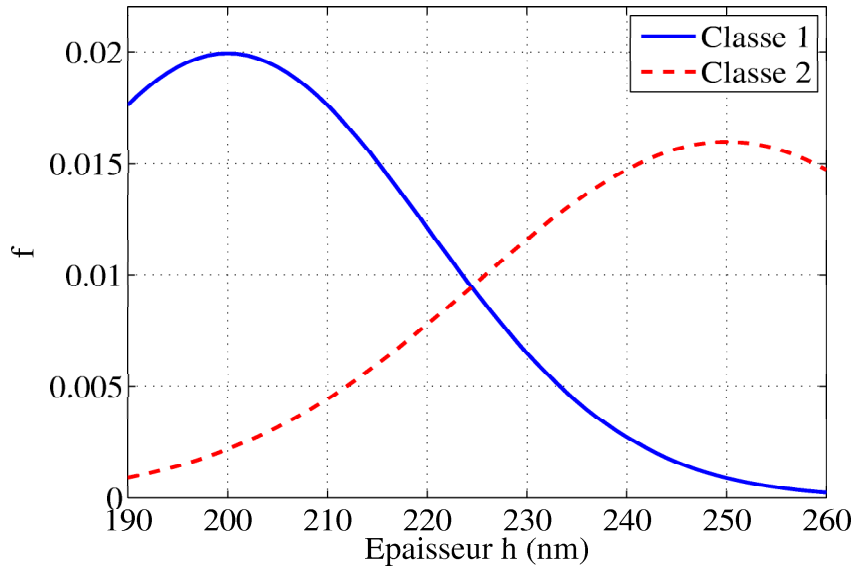


FIG. II.25: Densité de probabilité des épaisseurs pour les échantillons appartenant aux deux classes A et B.

Supposons également qu'on produise le même nombre d'échantillons pour les deux épaisseurs h_1 et h_2 , ce qui implique que $P_r(C_1) = P_r(C_2) = 1/2$. La formule de Bayes permet de déterminer directement les probabilités a posteriori $P_r(C_1|h)$ et $P_r(C_2|h)$ pour qu'un échantillon d'épaisseur quelconque h appartienne à l'une des deux classes, respectivement C_1 et C_2 :

$$P_r(C_1|h) = \frac{f(h|C_1)}{f(h|C_1) + f(h|C_2)} \quad (\text{II.41})$$

$$P_r(C_2|h) = 1 - P_r(C_1|h) \quad (\text{II.42})$$

La figure II.26 présente les probabilités a posteriori d'appartenance aux deux classes C_1 et C_2 en fonction de la valeur de h , calculées par la formule de Bayes. Ainsi un échantillon d'épaisseur 210 nm par exemple possède 80% de chance pour qu'il soit affecté à la classe C_1 et 20% à la classe C_2 . L'erreur de classement est le nombre d'échantillons d'épaisseur voisine de h_1 classés dans la classe C_2 et inversement. Elle correspond donc à la surface de recouvrement des deux courbes de la figure II.26.

Le classifieur de Bayes constitue un outil performant à condition que les densités de probabilité et les probabilités à priori soient connues exactement. Or, en pratique il est rare

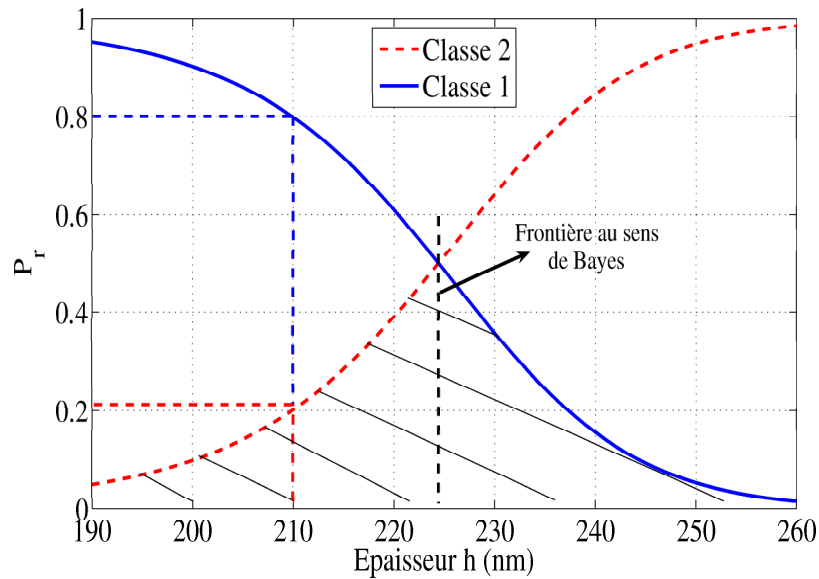


FIG. II.26: Probabilité *a posteriori* d'appartenance aux classes C_1 et C_2 en fonction de l'épaisseur h des échantillons.

que cette condition soit remplie. L'objet du paragraphe suivant est de montrer l'apport des réseaux de neurones dans ce contexte précis.

II.7.3 La voie neuronale

Dans un problème de classification, la propriété d'approximateur parcimonieux des RN est également mise à contribution. Soit Γ_m la variable aléatoire fonction de \vec{y} qui vaut 1 lorsque l'objet appartient à la classe m et 0 sinon. Le rôle du PMC est d'estimer en sortie les probabilités *a posteriori* pour qu'une entrée \vec{y} appartienne à une classe spécifique selon la nature du problème défini par un vecteur \vec{x} particulier.

$$\begin{aligned} s_m &= E[\Gamma_m | \vec{y}] = P_r(\Gamma_m = 1 | \vec{y}) \times 1 + P_r(\Gamma_m = 0 | \vec{y}) \times 0 \\ &= P_r(\Gamma_m = 1 | \vec{y}) \end{aligned} \quad (\text{II.43})$$

Le problème est alors ramené à un simple problème de régression non linéaire. Dans le cadre de notre travail, nous nous intéresserons plus particulièrement au cas de la classification dite supervisée, où la nature et le nombre de classes sont connus à l'avance.

Considérons le problème où l'on doit répartir les données à disposition en K classes distinctes. Dans le cas d'une classification, une classe doit être identifiée numériquement. En général, cette dernière est une information non numérique. Le RN ne pouvant pas être entraîné avec ce type de données, une représentation numérique doit être adoptée. La meilleure façon de représenter les classes est le codage binaire. Pour ce faire, chaque classe k est décrite par un vecteur de taille K , de telle sorte que la classe k , correspondant à un vecteur \vec{x} particulier soit codé par le vecteur : $[0_{(1)} \dots 1_{(k)} \dots 0_{(K)}]$.

L'architecture du RN dédiée à cette classification conserve les mêmes caractéristiques que celles du PMC (figure II.6). La dimension de la couche de sortie est égale à K . La seule modification notable concerne les neurones de sortie et plus particulièrement leurs fonctions d'activation. Le réseau étant destiné à estimer des probabilités, les valeurs des

neurones de sortie doivent être comprises dans l'intervalle $[0,1]$ et leur somme doit être égale à 1.

Il est préconisé [133] d'employer une fonction d'activation des neurones de sortie de type *softmax* (aussi connue en anglais sous le nom de fonction *multinomial logit*). La sortie m du réseau peut être écrite sous la forme suivante :

$$s_m = \frac{e^{E_m}}{\sum_{m=1}^K e^{E_m}} \quad (\text{II.44})$$

avec

$$E_m = \sum_{j=0}^{N_c} z_{m,j} O_j \quad (\text{II.45})$$

qui n'est autre que le potentiel du neurone m de sortie, et O_j est la valeur de sortie du neurone j dans la couche cachée donnée par :

$$O_j = g \left(\sum_{i=0}^{N_c} w_{j,i} x_i \right) \quad (\text{II.46})$$

La fonction mathématique softmax assure ainsi la normalisation des sorties dans l'intervalle $[0,1]$.

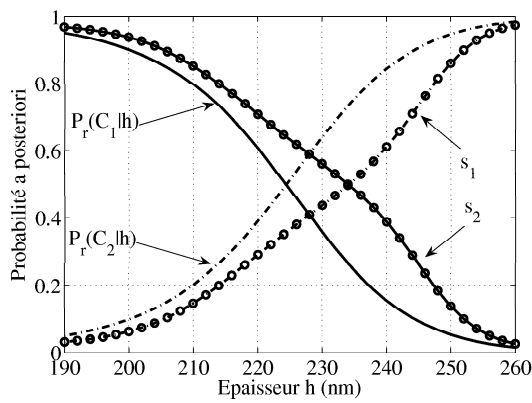
Nous insistons à ce stade sur le fait que le RN ne donne pas la classe de l'échantillon mais la probabilité a posteriori pour qu'une entrée appartienne à une classe donnée connaissant l'entrée \vec{y} . La décision d'affecter le résultat à une classe est ensuite prise par l'expérimentateur. Un moyen simple et efficace est d'utiliser la règle de décision de Bayes qui minimise les erreurs résultantes d'un mauvais classement. Mais, il se peut que les classes ne soient pas bien séparables pour un jeu de descripteurs donné. Ainsi, le pourcentage de mauvais classement se trouvera forcément élevé (figure II.26) quelles que soient les performances du classifieur. La seule solution consisterait à reconsidérer le choix des descripteurs.

La phase d'apprentissage d'un tel réseau classifieur est identique à celle des PMC décrite précédemment, à savoir l'ensemble des méthodes dites supervisées.

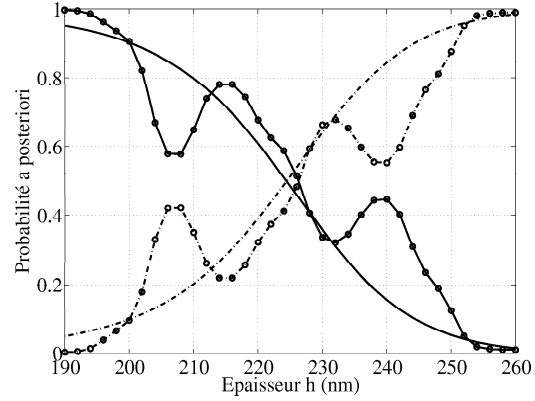
Reprenons le même exemple de classement d'échantillons de la partie précédente. Si le réseau de neurones est un bon estimateur de probabilités a posteriori, il devra retrouver le même résultat que celui donné par la règle de décision de Bayes sans connaissance a priori des densités.

Pour mettre en œuvre le classifieur neuronal, nous devons disposer d'un ensemble d'apprentissage {entrées, sorties}. Les entrées sont constituées de valeurs des épaisseurs et les sorties sont les classes correspondantes, codées en binaire. Les entrées sont générées aléatoirement selon une première loi normale de moyenne h_1 et d'écart type σ_{h_1} et correspondent à la classe C_1 , et une autre de moyenne h_2 et d'écart type σ_{h_2} pour les données de la classe C_2 conformément aux équations II.39 et II.40. Pour chaque entrée de la classe C_1 , on associe une sortie représentée par le vecteur $[1 \ 0]$, et un vecteur $[0 \ 1]$ pour celles de la classe C_2 . Il reste à préciser que pour l'apprentissage du classifieur, le nombre de couples employés dans chaque classe est égal.

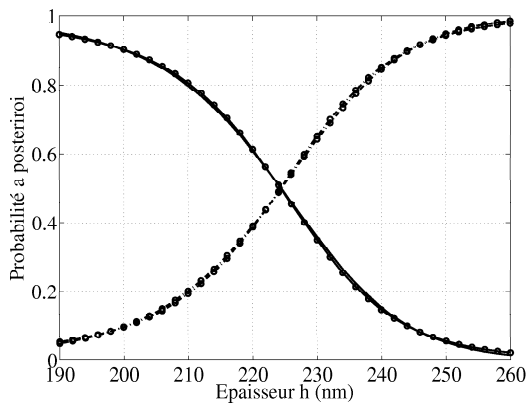
La figure II.27 montre les probabilités a posteriori (s_m) estimées par le RN pour dif-



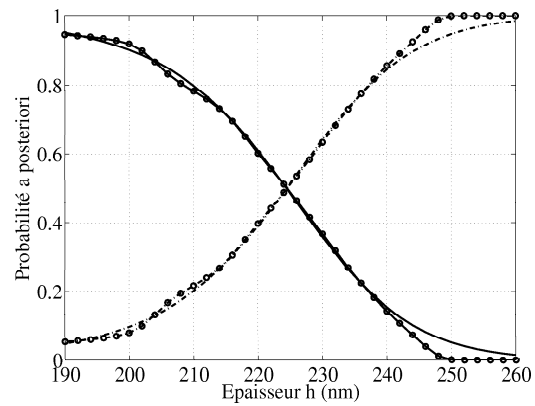
(a) 2 neurones, 200 échantillons



(b) 20 neurones, 200 échantillons



(c) 2 neurones, 4000 échantillons



(d) 20 neurones, 4000 échantillons

FIG. II.27: Probabilités a posteriori calculées par le réseau de neurones (traits+cercles) comparées à celles données par la règle de Bayes (traits continus) pour différents nombres de neurones et d'exemples d'entraînement.

férents neurones cachés et différents couples d'apprentissage. Les résultats sont comparés avec ceux donnés par la formule de Bayes (figure II.26).

Même si le problème est à une seule dimension, plusieurs enseignements sont à noter concernant la classification neuronale. Il est clair, d'après la figure II.27(a), que si le nombre d'exemples est petit, certaines régions de l'espace ne seront pas représentées et les probabilités estimées par le RN risquent d'être imprécises. L'augmentation du nombre de neurones ne sera pas bénéfique puisqu'on aura plus de degrés de liberté pour la fonction de régression mais toujours une mauvaise représentation de l'espace dû à l'insuffisance d'échantillons (figure II.27(b)). La solution consiste simplement à augmenter le nombre d'exemples pour balayer toutes les régions de l'espace des observations afin de modéliser efficacement la densité de probabilité. Dans les cas où il est difficile de générer un très grand nombre d'exemples dans l'ensemble d'apprentissage, il est souhaitable de réduire la dimension de l'espace des descripteurs.

D'après la figure II.27(c), où le nombre d'exemples utilisé est de 4000, les performances du classifieur sont meilleures même avec peu de neurones (ici 2) dans la couche cachée. Le réseau approche ainsi parfaitement la limite théorique donnée par la formule de Bayes

sans connaissance des densités II.39 et II.40. Il se peut qu'avec un nombre moins élevé d'exemples, les mêmes performances soient atteintes. En revanche, il faut être toujours prudent vis-à-vis du choix de nombre de neurones qui ne doit pas être exagérément large même si le nombre d'échantillons est suffisant. Le résultat peut en effet commencer à se dégrader (figure II.27(d)), car le RN va commencer à apprendre spécifiquement le corpus présenté. Il reste à noter que pour les problèmes ayant une topologie des classes plus complexe que celle présentée, ce qui est souvent le cas en pratique, un grand nombre d'exemples doit être pris dès le début pour représenter le mieux possible les différentes régions de l'espace des descripteurs.

Cet exemple assez simple nous a permis de comprendre le fonctionnement du RN destiné à résoudre un problème de classification. Par la suite, nous serons amené à mettre au point des classifieurs plus complexes dédiés à l'identification des signatures ellipsométriques, où le nombre de descripteurs, correspondant au nombre d'intensités mesurées vec_y , est beaucoup plus grand.

II.7.4 Application à l'identification des gammes d'épaisseurs de couches de résine

Dans cette partie, le classifieur sera appliqué au cas pratique de la caractérisation de couches minces. L'idée est illustrée par le schéma de la figure II.28. A partir d'une signature ellipsométrique mesurée \vec{y} sur un échantillon donné, dont on cherche à déterminer l'indice et l'épaisseur, un classifieur est mis au point dans le but d'identifier la classe d'épaisseur à laquelle appartient la couche d'indice inconnu constituant l'échantillon.

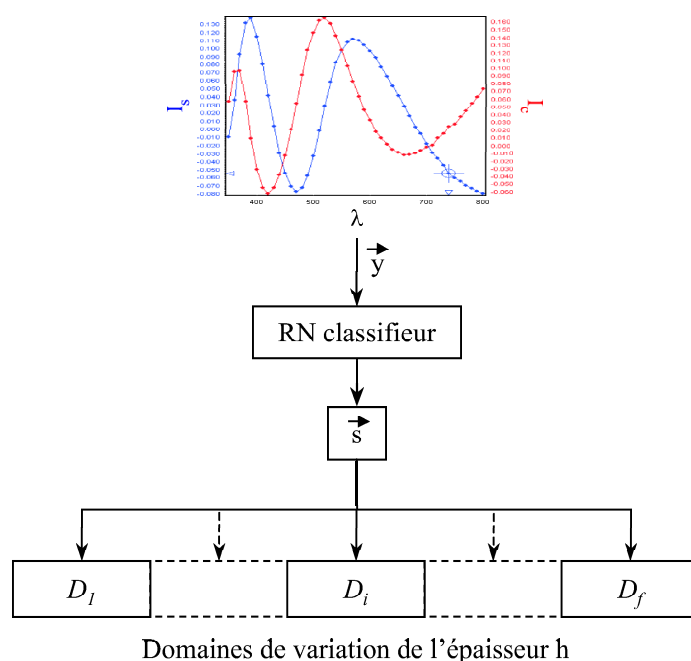


FIG. II.28: Schéma illustrant le principe de la classification neuronale appliquée à l'identification du domaine d'épaisseur d'une couche de résine déposée sur un substrat.

Le modèle structural utilisé pour réaliser cette étude est une couche mince de résine déposée sur un substrat de verre BK_7 (schéma de la figure II.1) d'indice supposé connu

(annexe A). Le classifieur est entraîné pour classer la signature ellipsométrique dans trois domaines d'épaisseurs différentes notées D_1 , D_2 et D_3 et données ci-dessous :

$$\begin{aligned} 200 \text{ nm} < D_1 < 400 \text{ nm} \\ 400 \text{ nm} < D_2 < 600 \text{ nm} \\ 600 \text{ nm} < D_3 < 800 \text{ nm} \end{aligned} \quad (\text{II.47})$$

Les signatures ellipsométriques employées pour l'entraînement du classifieur sont calculées à partir des paramètres tirés aléatoirement dans les domaines définis par II.47 pour l'épaisseur et II.23 pour l'indice de réfraction. L'entraînement est effectué avec des données bruitées comme il a été mentionné précédemment (paragraphe II.4.2). Afin d'éviter qu'une région de l'espace soit sous représentée, 3000 échantillons ont été choisis pour chaque classe. De même, le nombre de neurones est fixé à 5 afin d'assurer un bon niveau de généralisation. Les sorties du réseau, notées s_1 , s_2 et s_3 , sont codées binairement par un vecteur de dimension 3, à savoir $[1 \ 0 \ 0]$ pour les données qui appartiennent au domaine D_1 , $[0 \ 1 \ 0]$ pour D_2 et $[0 \ 0 \ 1]$ pour D_3 .

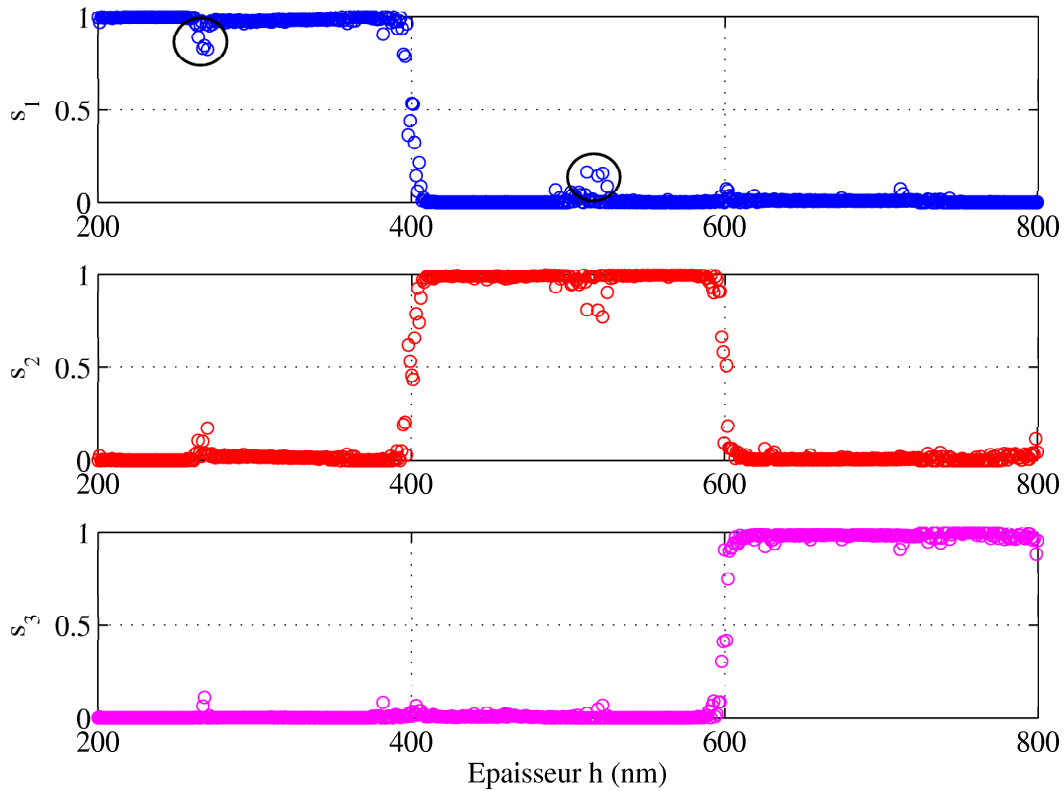


FIG. II.29: Sorties du réseau de neurones classifieur (s_1 , s_2 et s_3), représentant les probabilités *a posteriori* d'appartenance aux différentes classes.

Afin de tester le classifieur, nous avons généré une série d'échantillons dont l'épaisseur de la couche varie entre 200 nm et 800 nm par pas de 1 nm et l'indice est choisi aléatoirement dans le domaine d'entraînement (II.23). La figure II.29 montre les sorties (s_1 , s_2 et s_3) calculées par le classifieur. Rappelons que chaque sortie correspond à la probabilité *a posteriori* d'appartenance $P_r(C_k|\vec{y})$ à une classe k donnée ($k = 1, 2, 3$) connaissant les mesures \vec{y} . La somme des trois sorties pour un échantillon donné doit être égale à 1.

On remarque que pour les échantillons situés à proximité des frontières des différentes classes (400 nm et 600 nm), la probabilité pour que l'échantillon considéré appartienne à l'une ou l'autre des classes adjacentes est voisine de 0,5. En effet, la signature optique ne subit pas, a priori, de variation brusque autour de 400 nm et 600 nm, alors il est normal qu'un échantillon se trouvant à la frontière de 2 domaines soit classé à la fois dans l'un et dans l'autre en proportion de sa proximité. Notons que cette zone d'incertitude est relativement restreinte à moins de 10 nm.

Une solution peut être envisagée en mettant en œuvre un second classifieur tel que les classes englobent les frontières du premier. En revanche, le classifieur est très performant à l'intérieur de chaque classe et ne commet pas de mauvais classement dans les domaines d'épaisseurs étudiés si on applique la règle de décision de Bayes. Il faut également noter la présence, certes ici peu nombreuses, des sorties s déviant significativement de 0 ou de 1. Il s'agit essentiellement des échantillons d'épaisseurs voisines de 260 nm et de 500 nm (encadrés sur la figure II.29). Les signatures optiques de ces échantillons ont une légère tendance à être classées dans une catégorie différente de la leur. Ceci provient du fait qu'elles possèdent morphologiquement des signatures optiques légèrement différentes de leur classe respective. C'est typiquement la cause de présence de minima locaux. Il faut noter aussi que ces minima locaux dépendent fortement de la signature utilisée.

Les résultats théoriques validés, le RN classifieur peut désormais être appliqué expérimentalement. Pour cela, nous disposons de 3 échantillons, notés E_1 , E_2 et E_3 , dont l'épaisseur est estimée après fabrication respectivement voisine de 350 nm, 500 nm et 650 nm. Les signatures ellipsométriques mesurées sont introduites en entrée du classifieur défini précédemment. Le tableau II.10 regroupe les probabilités a posteriori estimées par ce dernier pour chacun des échantillons.

Sortie du classifieur \ Echantillons	E_1	E_2	E_3
s_1	0,999	0,005	0,008
s_2	0,001	0,995	0,005
s_3	0	0	0,987

TAB. II.10: Probabilités a posteriori estimées par le réseau de neurones classifieur à partir des mesures ellipsométriques effectuées sur les trois échantillons E_1 , E_2 et E_3 .

D'après les valeurs d'épaisseurs données, les résultats montrent que les échantillons se trouvent parfaitement bien classés. Nous devons souligner qu'il est tout de même possible de concevoir un classifieur fonctionnant sur une large zone de l'espace des épaisseurs, soit en élargissant chaque classe (ce qui serait souhaitable ici vu les bonnes performances) soit en augmentant le nombre de classes.

Rappelons que pour les méthodes d'optimisation classiques, il est souvent difficile de choisir le point de départ lorsque l'utilisateur ne possède pas une idée sur l'échantillon à caractériser. La classification peut dans ce cas être couplée à ces méthodes pour faciliter la recherche de la solution en diminuant fortement le risque de convergence prématurée. En effet, la présence des minima locaux est le problème majeur de ces méthodes. A titre d'exemple, la figure II.30 montre les valeurs de l'épaisseur calculées par la méthode de LM pour l'échantillon E_1 en fonction du point de départ de l'optimisation. Cette courbe confirme bien la présence d'un minimum local illustrée par la figure II.29, pour

une épaisseur de l'ordre de 260 nm. Au delà de ce point, la valeur de l'épaisseur est de l'ordre de 364 nm, conformément à celle donnée dans le tableau II.6. Le problème de la présence de minima locaux est omniprésent même dans ce cas extrêmement simple. Il est encore plus difficilement résolu lorsque la complexité du problème est grandissante.

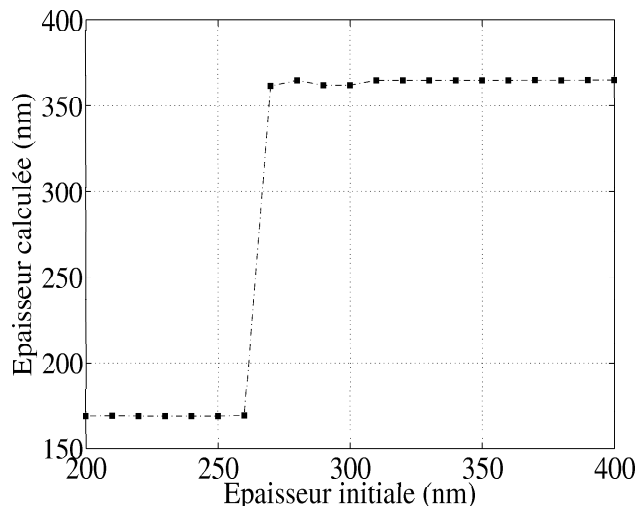


FIG. II.30: Valeurs de l'épaisseur h calculées par optimisation classique (algorithme de Levenberg-Marquardt) en fonction de l'épaisseur initiale h_0 (point de départ) dans le cas de l'échantillon E_1 .

La classification présentée ici permet de localiser la solution dans l'espace des paramètres. Ce résultat n'est pas anodin et peut conduire à de fortes améliorations de la caractérisation neuronale. En effet, la connaissance qualitative de l'épaisseur va permettre de restreindre le domaine d'entraînement du réseau de neurones qui sera ensuite utilisé pour la caractérisation proprement dite. Dans ce cas, les performances du réseau sont meilleures sans avoir besoin d'un très grand nombre d'exemples d'entraînement ou de nombre de neurones.

II.8 Conclusion

Ce chapitre était consacré à la présentation des RN, en particulier les PMC qui constituent de véritables approximateurs universels parcimonieux. Nous avons ainsi détaillé leur mode de fonctionnement et tout particulièrement le rôle de la phase d'apprentissage nécessaire avant toute étape d'exploitation du réseau.

Les potentialités du RN à résoudre le problème inverse en ellipsométrie ont été étudiées et appliquées à la caractérisation de couches minces de résine. En effet, le RN est capable d'établir la relation qui existe entre ses entrées (signature optique mesurée) et ses sorties (paramètres optiques et géométriques de la couche de résine). Comme le RN fonctionne sur un domaine limité de l'espace des paramètres, ses performances ont été étudiées en fonction de la largeur de ce dernier. Cependant, il faut toujours s'assurer que le nombre d'exemples d'entraînement soit suffisant pour représenter tout l'espace choisi, tout en gardant un bon compromis entre ce dernier et le nombre des neurones cachés. Nous avons ensuite vu que le RN constitue également un bon outil permettant la détermination des intervalles de confiance des paramètres recherchés.

Notre attention s'est ensuite portée sur la limitation du temps global de caractérisation, à savoir le temps de mesure suivi du temps de traitement. Le temps de traitement par RN étant quasi instantané, nous nous sommes concentrés sur le temps de mesure. Ainsi, nous avons présenté une méthode (la méthode HVS) pour sélectionner les mesures pertinentes au RN, ainsi que pour choisir une architecture simplifiée en terme de neurones cachés tout en gardant de bonnes performances. Ceci possède des avantages aussi bien au niveau expérimental que neuronal. En effet, le temps de mesure se trouve considérablement réduit par la suppression des mesures non pertinentes. De plus, la diminution du nombre de paramètres à optimiser au sein du RN permet de réduire les problèmes liés aux minima locaux lors de la phase d'entraînement et conduit ainsi à de bonnes performances en généralisation.

Enfin, dans la dernière partie de ce chapitre, nous avons fourni une autre application des RN : la classification. Dans le cadre du travail présenté, nous avons employé cette fonctionnalité pour identifier la gamme d'épaisseur de couches minces de résine avant la phase de caractérisation. Cette technique peut également être couplée à d'autres méthodes d'optimisation classique, dans le but d'aider à choisir efficacement le point de départ pour éviter que l'algorithme soit piégé trop rapidement dans un minimum local. Cela ouvre un large champ d'applications potentielles comme par exemple la détection de rugosité, la présence ou non d'une couche quelconque, le choix du profil géométrique de la structure à étudier, etc. . .

Le chapitre suivant est entièrement consacré à la caractérisation neuronale dans le cadre précis de la scatterométrie.

CHAPITRE III

Caractérisation des réseaux de diffraction

Ce troisième et dernier chapitre expose et analyse les différentes contributions des réseaux de neurones dans le domaine de la scatterométrie. Les structures diffractantes étudiées visent des applications en optique et en microélectronique et sont ainsi gravées sur différents types de substrats avec des profils géométriques variés. La période sera généralement supposée connue, sauf dans certains cas où l'on sera amené à la déterminer.

La première partie a pour but de présenter au lecteur les enjeux liés à la réalisation des réseaux optiques en insistant en premier lieu sur les différentes techniques de fabrication. Ensuite, nous intéresserons à l'aspect métrologique permettant la caractérisation dimensionnelle de ces structures.

La seconde partie concerne l'application des RN pour la résolution du problème inverse en scatterométrie. Elle concerne la détermination des paramètres définissant le profil géométrique d'un réseau de diffraction. Cette étude est ensuite validée sur un réseau de diffraction gravé sur un substrat de silice. Les résultats obtenus sont comparés avec ceux donnés par deux techniques microscopiques complémentaires, à savoir la microscopie à force atomique et la microscopie électronique à balayage. Pour des raisons techniques, la période de ce réseau est donnée avec un certain degré d'incertitude. Nous allons voir comment ce paramètre peut être facilement déterminé par voie neuronale. L'évaluation du temps de traitement par RN sera également abordée. Nous démontrerons ensuite la faisabilité de la méthode neuronale pour la caractérisation des réseaux de diffraction possédant différents profils géométriques de complexités différentes. Puis, nous finirons par donner une application directe : le contrôle de l'homogénéité d'une structure.

La troisième partie traite d'une étude paramétrique détaillée permettant d'évaluer le comportement du RN en fonction de différents paramètres. Cette étude concerne d'une part les paramètres propres au RN, et d'autre part les paramètres qualifiés d'externes liés essentiellement aux conditions expérimentales de mesure ainsi qu'aux hypothèses structurelles.

Puis, nous appliquerons la méthode HVS permettant de réduire le nombre de mesures nécessaires pour garantir une caractérisation optimale. Cela a pour avantage de diminuer le temps global de caractérisation lié essentiellement au temps nécessaire à la mesure de la signature ellipsométrique utilisée. L'application immédiate concerne un contrôle plus

rapide de l'homogénéité d'une structure.

Enfin, dans la dernière partie du chapitre, nous nous intéresserons aux différentes contributions de la classification par voie neuronale en scatterométrie, en particulier l'identification du profil géométrique de l'échantillon sous test.

III.1 Les enjeux de la scatterométrie

Nous commencerons ce paragraphe par donner au lecteur une idée sur l'avancement technologique de la fabrication de structures avec des dimensions de plus en plus réduites, dans le cadre d'application en microélectronique ou en optique. Nous présenterons ensuite les principales méthodes de fabrication utilisées dans ce contexte. Puis, nous nous intéresserons au volet de la caractérisation de ces structures, portant sur la mesure dimensionnelle, en positionnant les méthodes optiques par rapport à celles communément utilisées.

III.1.1 Besoins des futurs nœuds technologiques

La miniaturisation est un terme qui subsiste toujours et qui gouverne le marché économique de la technologie. Il est indéniable que l'acteur principal dans l'évolution technologique, depuis quelques années jusqu'à aujourd'hui, est l'industrie de la microélectronique. Lorsque l'on traite du sujet, on se réfère inévitablement à la "loi de Moore" [134] décrivant la vitesse d'évolution de la densité d'intégration des circuits intégrés. Cette loi est étonnamment vérifiée jusqu'à l'heure actuelle et continuera sûrement de l'être jusqu'en 2015 d'après les prédictions des scientifiques. Pour atteindre ce but, l'industrie de la microélectronique continue d'explorer la limite ultime de ce qui est possible. La figure III.1 permet de chiffrer cette évolution en terme de dimension critique (ou CD pour *Critical Dimension*) à atteindre dans les années à venir, comme il a été prévu par la communauté scientifique ITRS (*International Technology Roadmaps for Semiconductor*) en 2004.

Remarque : Le CD désigne la plus petite dimension que l'on peut obtenir dans un processus de fabrication donné. Dans l'industrie de la microélectronique, ce terme définit un nœud ou une génération technologique.

D'autres domaines comme l'optique, connaissant un véritable essor à l'aube de ce siècle, cherchent à transférer au sein de leurs activités le savoir faire de l'industrie de la microélectronique dans le but de réaliser des composants nanostructurés. Le contrôle des dimensions caractéristiques d'un composant donné est un point clef pour sa réussite. En microélectronique comme en optique cette étape est essentielle pour assurer les propriétés requises pour le dispositif. Ce contrôle nécessite avant tout une maîtrise des techniques de fabrication. Puis, le résultat doit être vérifié à l'aide de méthodes de caractérisation fiables et précises. Cette seconde étape vise plus particulièrement les dimensions géométriques qui définissent la fonction qui sera attribuée au dispositif. Elle entre ainsi dans le cadre de la métrologie dimensionnelle.

Bien que la fabrication et la caractérisation soient deux étapes primordiales, les fonds investis dans l'une ou dans l'autre aujourd'hui ne sont pas semblables. C'est sans doute la première qui l'emporte, ce qui a favorisé l'écart au niveau de développement de chacune

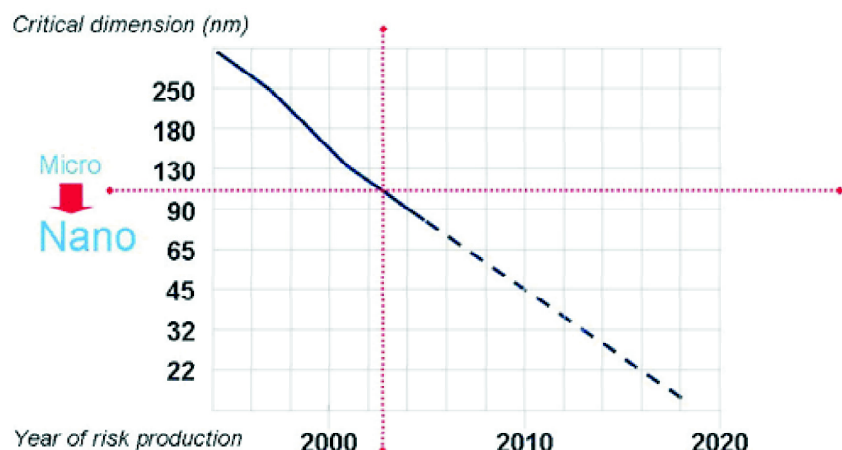


FIG. III.1: Evolution de la fabrication selon la loi de Moore connue en microélectronique selon les prévisions de l'ITRS (International Technology Roadmaps for Semiconductor) en 2004.

d'elles. Quant au milieu industriel, il ne s'est pas soucié de ce fait puisque les méthodes relatives à la métrologie dimensionnelle dont il dispose étaient, jusqu'à présent, suffisantes et pouvaient répondre aux critères recherchés. Or, dans les dernières années, la réduction continue du CD a révélé le besoin de techniques de caractérisation de plus en plus performantes et facilement intégrables en ligne.

III.1.2 Procédés de fabrication

L'augmentation continue de la densité d'intégration des composants optiques et électroniques exigée par le milieu industriel, est essentiellement régie par les techniques de fabrication utilisées. Dans cette section, les techniques de lithographie les plus utilisées actuellement pour réaliser des structures submicroniques seront présentées.

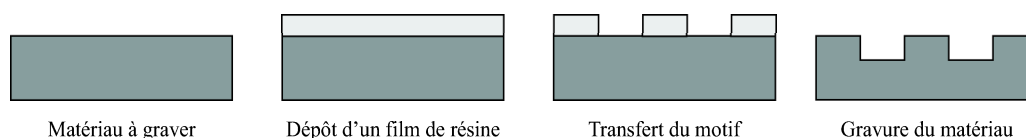


FIG. III.2: Etapes de gravure d'un matériau à travers un masque de résine.

Nous nous intéresserons plus particulièrement à celles qui consistent à transférer d'abord un masque de résine sur le matériau à graver, suivi d'une gravure à travers ce masque afin d'obtenir la structure souhaitée. Une représentation schématique des différentes étapes technologiques est montrée sur la figure III.2. Nous aborderons également l'influence de la technique de fabrication sur la forme du profil géométrique final obtenu.

La lithographie optique a été le facteur clef pour réduire la taille de motifs dans l'industrie de la microélectronique. Elle consiste à reproduire un motif par insolation (rayonnement UV) à travers un masque sur une résine photosensible (figure III.3). L'interaction des rayons avec la résine modifie ses propriétés de solubilité. Ainsi, après révélation dans un développeur (solution basique), la zone exposée va subsister ou être éliminée selon que la résine soit respectivement de tonalité négative ou positive. Le motif reproduit n'est

autre que l'image du masque à travers un système optique particulier, ce qui permet de réduire les dimensions du motif $\times 4$ ou $\times 5$ par rapport à celles du masque.

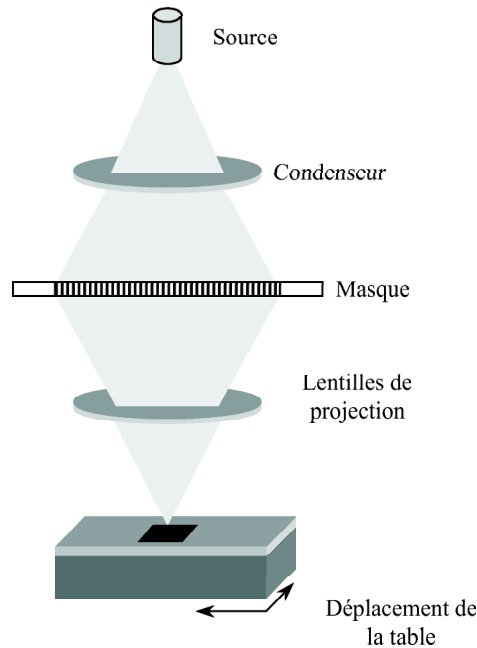


FIG. III.3: Principe de la lithographie optique par projection.

La résolution, notée R , offerte par la lithographie optique peut être dérivée de l'équation de Rayleigh de la façon suivante :

$$R = k_1 \frac{\lambda}{ON} \quad (\text{III.1})$$

avec λ est la longueur d'onde utilisée, ON est l'ouverture numérique du système optique et k_1 est une constante positive introduite pour représenter la qualité de la résine, mais également les éléments du système optique et le masque. La lithographie optique utilisant des longueurs d'onde de 248 nm et 193 nm permet de fabriquer des lignes de largeur 45 nm.

On peut citer également la technique de fabrication de masque par holographie [135]. Cette dernière consiste à faire interférer deux ondes lumineuses sur la surface de la résine photosensible. La forme du profil obtenu est sinusoïdale, mais il est difficile d'obtenir des réseaux qui débouchent sur le substrat.

Il existe d'autres techniques émergentes de la lithographie optique : la lithographie à immersion [136] dont le but est d'améliorer l'ouverture numérique du système optique en utilisant un fluide convenable entre les lentilles de projection et l'échantillon, la lithographie double exposition [137] et la lithographie extrême ultraviolet [138] travaillant à une longueur d'onde de 13,4 nm. Ces techniques suscitent l'attention de l'industrie de la microélectronique pour la réalisation des nœuds technologiques futurs.

La lithographie électronique est une méthode dite directe [139, 140]. A l'aide d'un faisceau d'électrons balayant la surface, les motifs sont formés sur une résine électrosensible préalablement déposée sur le substrat à graver, sans l'utilisation d'un masque

(figure III.4). Il faut cependant noter que la résine doit déboucher sur le substrat pour une gravure ultérieure optimale, en réglant convenablement l'intensité du faisceau d'électrons incident.

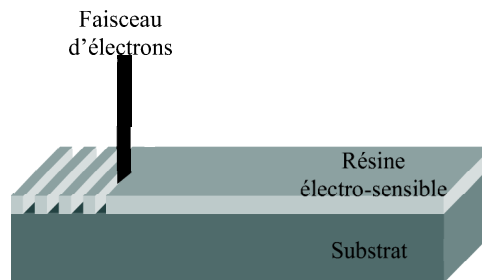


FIG. III.4: Principe de la lithographie électronique : écriture directe par un faisceau d'électrons sur une couche de résine électro-sensible.

Cette technique est une amélioration de la lithographie par faisceau optique (dont la taille du spot est environ 100 fois plus grande). Elle est limitée par la rétrodiffusion des électrons secondaires par le substrat vers la résine. Cet *effet de proximité* peut conduire à une insolation de zones de résine non désirées et élargir ainsi les dimensions du motif réalisé.

Les dimensions obtenues sont conditionnées par la taille du faisceau électronique avec une résolution spatiale de l'ordre de 20 nm. Ainsi, cette technique permet l'obtention de motifs de très faible période ayant des flancs presque verticaux (proche de 90°), homogènes sur de grandes surfaces à condition que le dispositif de gravure soit bien aligné. Cependant, du fait de sa lenteur d'écriture, la lithographie électronique directe n'est jamais apparue comme une solution crédible pour la production en série.

La lithographie par nanoimpression (NIL *NanoImprint Lithography*) est une technique simple, apparue récemment [141] pour la fabrication de structures nanométriques. Le principe consiste à dupliquer des motifs contenus sur un moule dans une couche de résine déposée sur un substrat à graver (figure III.5). Le moule étant fabriqué par les techniques de lithographie standard. On distingue deux grands types de nanoimpression : la nanoimpression thermique et la nanoimpression assistée par ultraviolet (UV-NIL *Ultra Violet curing NanoImprint Lithography*).

Dans la première, un moule contenant les motifs à reproduire est pressé dans le polymère chauffé au delà de sa température de transition visqueuse. A cette température, le polymère devient visqueux et se déforme facilement. Après refroidissement, le moule est retiré et les motifs sont transférés dans le polymère. Afin d'éviter tout contact entre le moule et le substrat, une couche résiduelle de polymère est laissée au fond des motifs. Cette dernière est ensuite éliminée par gravure plasma.

Dans la seconde, la résine est exposée à un rayonnement ultraviolet après pressage, à travers le moule qui est transparent. Ceci entraîne la photo-polymérisation de la résine et le moule peut être retiré. Les étapes suivantes de gravure sont les mêmes que dans la nanoimpression thermique. Les lecteurs intéressés par cette technique trouveront plus de détails dans la référence [142].

L'avantage de cette technique est qu'elle n'est pas limitée par la diffraction, la diffusion ou l'interférence de l'onde dans le polymère vu qu'elle n'utilise aucun faisceau énergétique. La forme du profil transféré dans la résine est directement liée à la qualité et la forme

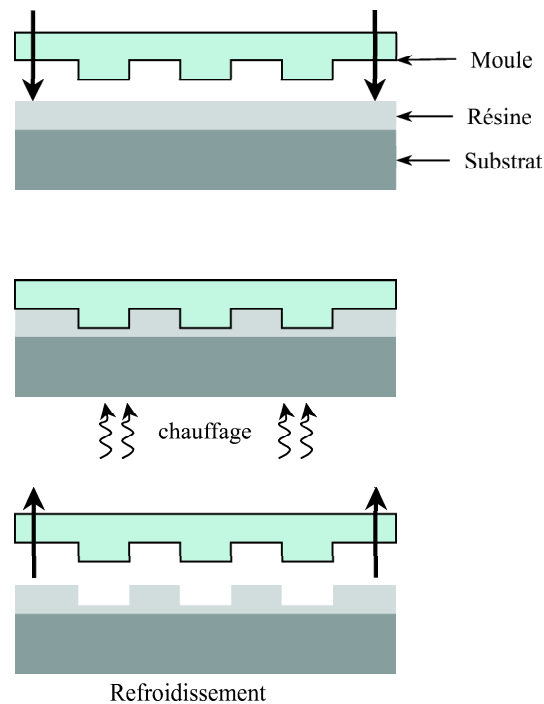


FIG. III.5: Principe de la lithographie par nanoimpression thermique.

du moule utilisé. En général, comme un seul moule est suffisant pour la réalisation de plusieurs impressions, ce dernier est fabriqué dans des conditions de précision optimales. Si le processus d'impression se déroule correctement, le masque est parfaitement imprimé. Dans le cas contraire, où par exemple les cavités du moule ne sont pas remplies, on peut obtenir des profils qui peuvent évoluer d'une forme sinusoïdale (au début du pressage) jusqu'à des profils présentant des arrondis (ce point sera illustré ultérieurement dans les résultats du chapitre).

Cette technique constitue une voie prometteuse pour la réalisation des futurs nœuds technologiques avec une résolution de l'ordre de 5 nm. Sa principale limitation provient du moule qui, entrant en contact avec la résine un grand nombre de fois, s'use et devient difficile à nettoyer.

Après avoir présenté les principales techniques permettant la structuration de la résine en tant que masque, nous allons à présent décrire les étapes de gravure permettant de transférer le motif dans le substrat.

Les techniques de gravure permettant le transfert du masque de résine, décrit par les précédentes techniques, dans le substrat sont communes à toutes les techniques de lithographie présentées. On peut distinguer deux cas : la gravure *sélective* et la gravure *non sélective*. Dans la première, le matériau est préférentiellement gravé, c'est donc la méthode de gravure (isotrope ou anisotrope) utilisée qui conditionne la forme du profil final. Dans la deuxième, le matériau et la résine se gravent à la même vitesse. Le profil final, dans ce cas, dépend fortement de la forme du masque.

Deux voies sont généralement possibles en gravure : la gravure humide et la gravure sèche.

La gravure humide consiste à tremper le substrat contenant le masque dans une solution chimique. Les zones non protégées de la résine subissent une réaction chimique et seront éliminées. L'inconvénient majeur de ce type de gravure est sa nature isotrope. En effet, la solution chimique peut agir latéralement et engendrer une déformation des flancs du motif (figure III.6).

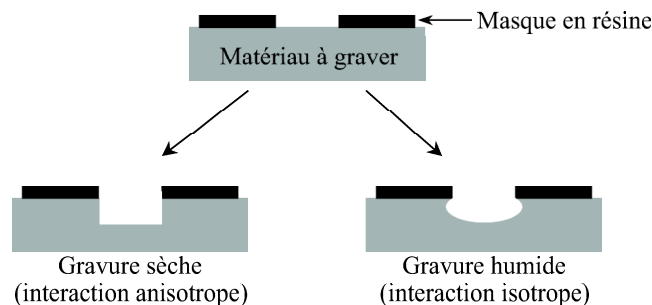


FIG. III.6: Influence de la gravure sèche anisotrope et humide isotrope sur la forme du profil final.

La gravure sèche utilise un faisceau d'électrons en présence d'un plasma pour graver le matériau. Elle est la plus utilisée pour la fabrication des réseaux de faibles périodes. Un faisceau d'électrons permet de graver le matériau en présence d'un plasma. L'interaction avec le matériau est anisotrope ce qui limite la déformation du profil géométrique obtenu. On définit trois méthodes de gravure selon la nature de l'attaque :

- *IBE* (en anglais *Ion Beam Etching*) qui est de nature physique. Le faisceau ne comporte pas d'éléments qui interagissent chimiquement avec le matériau à graver.
- *RIBE* (en anglais *Reactive Ion Beam Etching*) qui est de nature physico-chimique. Le bombardement des ions ainsi que la réaction chimique du plasma avec le matériau conduisent à l'arrachement de la matière.
- *RIE* (en anglais *Reactive Ion Etching*) qui est de nature chimique. Dans ce cas, le faisceau d'ions apporte l'énergie nécessaire au processus pour que la gravure puisse avoir lieu.

Nous avons vu que la forme du profil géométrique final dépend de la technique de fabrication utilisée. Il est ainsi important de comprendre le déroulement de ces étapes de fabrication afin de prendre en compte d'éventuels défauts lors de la modélisation géométrique ultérieure de la structure.

III.1.3 Moyens de caractérisation

La dernière opération dans une chaîne de fabrication d'un composant est l'étape de caractérisation. Elle permet la validation ou non de la structure ainsi réalisée. Cette section porte essentiellement sur les méthodes largement utilisées dans le cadre de la mesure dimensionnelle de structures périodiques.

III.1.3.1 Méthodes microscopiques

Le **microscope électronique à balayage MEB** est considéré comme l'outil de base dans l'industrie de la microélectronique pour la caractérisation des structures de

faible dimension. Son utilisation est devenue très courante dans la plupart des domaines scientifiques comme en biologie, chimie, médecine, métallurgie, etc . . .

Le principe de fonctionnement du MEB est basé sur l'interaction entre un faisceau d'électrons et un matériau constituant la surface à analyser. Cette interaction produit une variété de collisions élastiques et inélastiques entre les électrons et les atomes au sein de l'échantillon. Les différents types de rayonnements issus de cette interaction (électrons secondaires, électrons retrodiffusés, électrons Auger, rayons X, etc. . .) révèlent des informations exploitables à propos de l'échantillon incluant la morphologie externe (texture), la composition chimique, la structure cristalline, etc . . .

Les deux types de MEB qui sont largement employés pour la mesure des dimensions critiques sont le **X-SEM** et le **CD-SEM** (de l'anglais *Critical Dimension Scanning Electron Microscope*). Dans le premier, l'échantillon est clivé et une vue de coupe peut être réalisée sous différents angles. Ceci nous permet d'observer la morphologie des profils à condition que l'observation soit effectuée parallèlement aux sillons du réseau. L'inconvénient dans ce cas réside dans l'aspect destructif de la caractérisation. Dans la deuxième, l'observation est réalisée par vue de dessus. Cette technique constitue le moyen de vérification de la largeur des lignes fabriquées (ou CD). En revanche, elle ne permet pas d'accéder à la forme du profil.

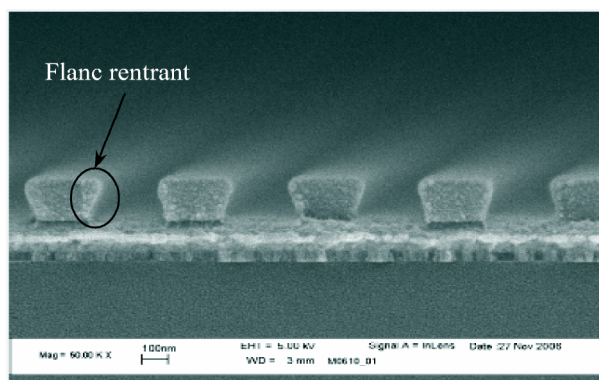


FIG. III.7: Image MEB d'un réseau en résine déposé sur un substrat d'or (Source : Laboratoire Hubert Curien, Saint-Etienne).

Bien que la caractérisation offerte par le MEB reste très locale, ce dernier est un outil qui permet de fournir une image non déformée de la structure sans connaissance a priori. Il est ainsi capable de détecter les défauts souvent difficiles à observer avec d'autres techniques. A titre d'exemple, la figure III.7 montre, selon une vue de coupe, un réseau en résine déposé sur un substrat d'or dont le profil présente des flancs rentrants. Par ailleurs, l'image obtenue par MEB doit être traitée pour remonter aux paramètres géométriques du profil. Cette étape nécessite l'utilisation des algorithmes de traitement d'image assez sophistiqués à ne pas négliger pour obtenir des résultats corrects. Il faut noter aussi que le faisceau d'électrons peut causer un endommagement du profil, d'où la nécessité d'un traitement préalable particulier de l'échantillon (par exemple métallisation).

La **microscopie à force atomique** ou **AFM** (de l'anglais *Atomic Force Microscopy*) est considérée comme une évolution de la microscopie à effet tunnel (*Scanning Tunneling Microscope* ou STM). Cette technique a été introduite par Binnig et ses collaborateurs en 1986. Bien que la résolution du STM puisse atteindre le niveau moléculaire ou atomique, il est impossible d'observer un matériau isolant avec cette dernière. Ce manque a motivé

ce même groupe à développer l'AFM permettant ainsi de réaliser des topographies sur des échantillons non conducteurs.

Depuis sa mise au point, cette technique a connu un essor fulgurant et s'est imposée comme un outil incontournable de caractérisation des surfaces permettant d'obtenir des images tridimensionnelles à l'échelle nanométrique indépendamment de la nature des échantillons.

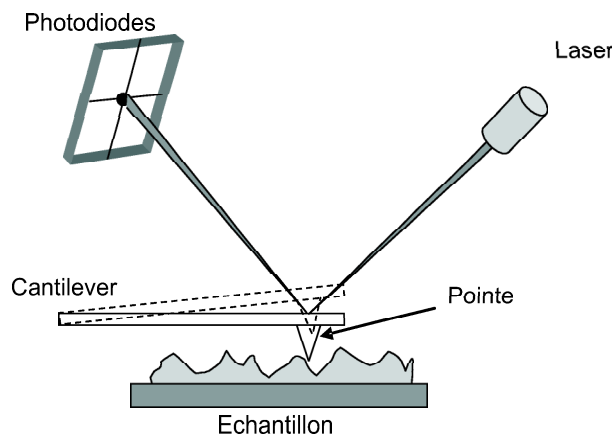


FIG. III.8: Principe de fonctionnement du microscope à force atomique.

L'AFM appartient à la famille des microscopes à sonde locale. Le principe (figure III.8) consiste à sonder la surface d'un échantillon à travers une pointe (ou *tip* en anglais). Elle doit être particulièrement fine et sa composition chimique doit lui conférer des propriétés de dureté évidentes. Cette pointe est montée perpendiculairement à l'extrémité d'un bras de microlevier élastique, appelé *cantilever* qui, à son tour, doit avoir une grande fréquence de résonance et une faible raideur. Ce dernier peut être déplacé finement par rapport à l'échantillon grâce à un système de tubes piézo-électriques dans les trois directions de l'espace. Les forces attractives ou répulsives entre la surface et la pointe induisent la déflexion du levier. Afin de mesurer cette déflexion tout au long du balayage, un laser est focalisé à l'extrémité du levier juste au dessus de la pointe et le faisceau réfléchi est envoyé sur un système de photodiodes. La formation d'une image est possible grâce au balayage du levier ligne par ligne sur la surface de l'échantillon.

La pointe est l'élément essentiel de cette technique. Les pointes classiques sont la plupart du temps obtenues par les mêmes techniques de gravure utilisées en microélectronique. Le challenge réside dans la fabrication des pointes de plus en plus fines tout en gardant les mêmes propriétés (duretés, flexibilité, ...). La forme du profil mesuré est donc directement liée à la forme et à la taille de la pointe utilisée (figure III.9). En effet, le profil mesuré n'est autre que la convolution de la pointe avec la forme réelle des motifs. Les scientifiques continuent à faire évoluer cette technique en introduisant des pointes de plus en plus sophistiquées (par exemple en nanotubes de carbone).

L'AFM est une méthode qui fournit une image de la surface analysée, ne nécessitant pas d'hypothèse préalable sur le profil géométrique. Par contre, de la même façon que pour le MEB cette image devra être traitée afin d'extraire la forme du profil mesuré. Cette technique ne permet qu'une caractérisation locale qui s'étend juste sur quelques périodes, et devient très longue dans le cas d'analyse de larges surfaces. Contrairement au MEB, l'AFM permet une caractérisation non destructive.

Elle est cependant utilisée comme une alternative au SEM et au TEM [143].

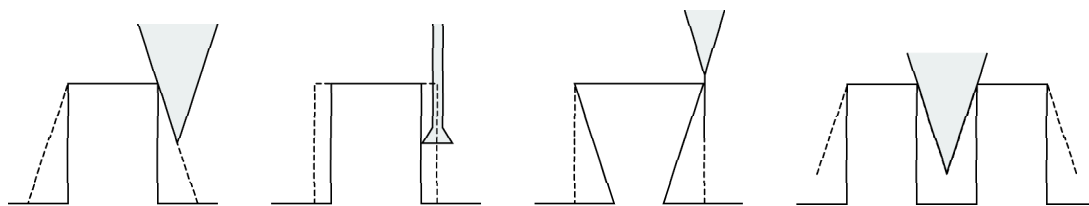


FIG. III.9: Influence de la forme de la pointe AFM sur la mesure du profil géométrique. Les traits pointillés et continus désignent respectivement le profil mesuré et le profil réel.

Il existe une variante de l'AFM, appelée CD-AFM, apparue récemment pour la caractérisation de la rugosité des lignes du réseau [144]. Cette dernière utilise une pointe évasée permettant de scruter les flancs des sillons.

Malgré les avantages que possèdent les deux méthodes microscopiques présentées, elles commencent à atteindre leurs limites de résolution spatiale. Les futurs nœuds technologiques envisagés visent donc à développer des techniques alternatives permettant d'atteindre les performances souhaitées.

III.1.3.2 Méthodes indirectes

L'émergence des méthodes optiques indirectes à caractère non destructif a redonné l'espoir aux expérimentateurs de pouvoir pallier les limitations des méthodes classiques que l'on vient de présenter.

La scatterométrie est une méthode générale qui, à l'heure actuelle est presque exclusivement utilisée dans le cas de structures périodiques diffractantes. Le principe est basé sur deux étapes bien distinctes (figure III.10) :

- Mesure d'une signature optique diffractée par la structure périodique dans des conditions expérimentales données,
- Détermination indirecte du profil géométrique de la structure par résolution d'un problème inverse sous certaines hypothèses préalables (forme du profil, indices connus, etc...).

L'investissement dans le développement et l'application de cette technique pour la caractérisation des structures périodiques a commencé d'être notable au début des années 90 [145, 146, 147, 148]. D'après les prévisions de l'ITRS en 2007 [149] (figure III.11), la scatterométrie a été sélectionnée parmi les solutions potentielles pour la métrologie dimensionnelle en lithographie jusqu'en 2016 où le CD est prévu entre 22 nm et 16 nm, ce qui a été mis en évidence également par une équipe de recherche de la société Intel [150].

La signature optique employée en scatterométrie peut être de différentes natures provenant de divers types d'instrument optique. On peut citer les systèmes permettant de mesurer l'intensité diffractée en fonction de l'angle d'incidence θ [151] (appelée également 2θ -scatterométrie), de la longueur d'onde λ (scatterométrie spectroscopique) [152] ou même de l'angle d'azimut [153] généralement dans l'ordre de diffraction spéculaire. Les ordres supérieures en 2θ -scatterométrie lorsqu'ils existent ont été également exploités [151]. La scatterométrie utilisant une signature ellipsométrique spectroscopique est largement plus étudiée [154, 155]. Récemment, Novikova et al. ont mis en évidence les potentialités de l'ellipsométrie de Mueller en diffraction conique [156]. Ils ont montré que ce type de signature permet une meilleure sensibilité aux profils géométriques en réduisant les corrélations entre ses différents paramètres. Parmi les instruments développés récem-

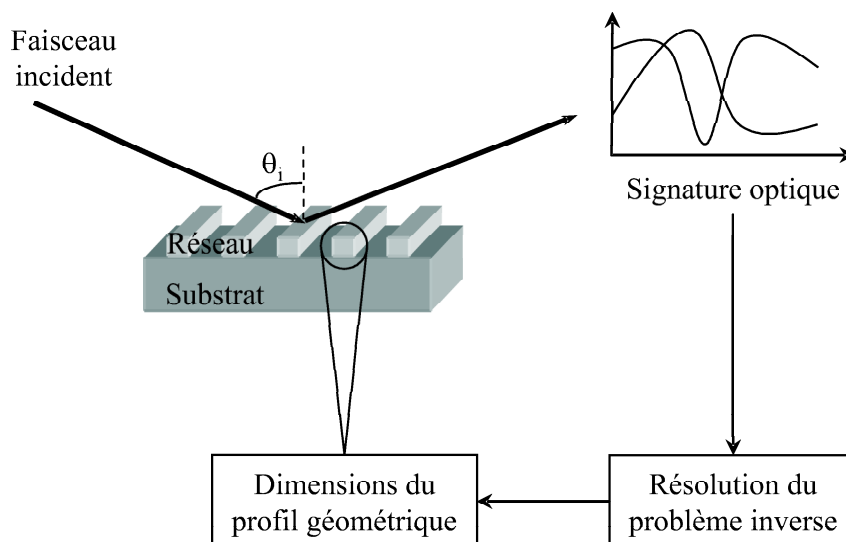


FIG. III.10: Principe de la scatterométrie : à partir de la signature optique mesurée sur une structure périodique, les dimensions du profil géométrique sont caractérisées par résolution d'un problème inverse.

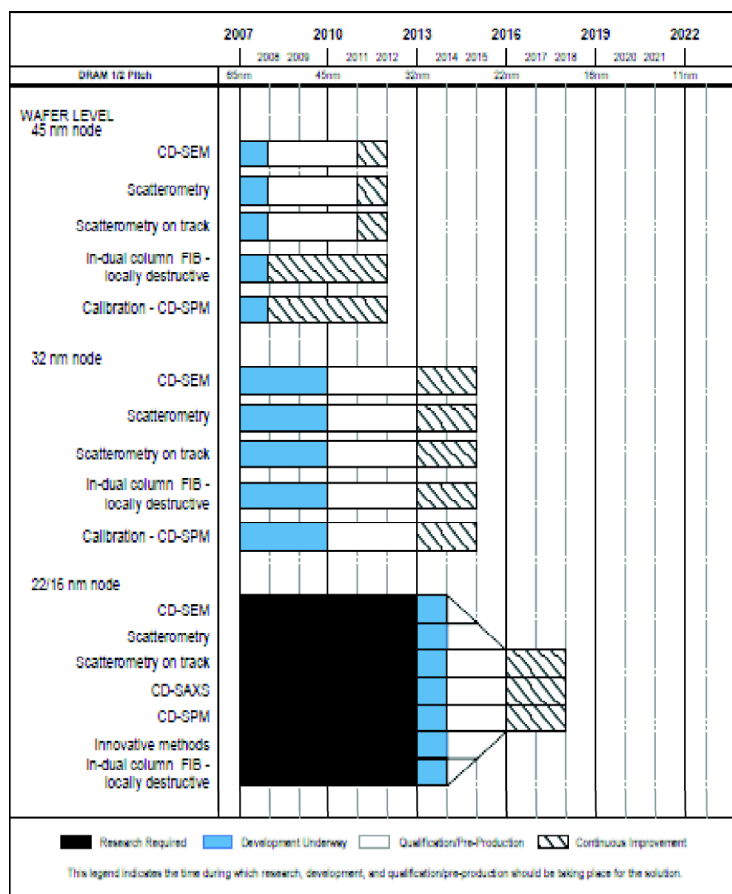


FIG. III.11: Solutions potentielles envisagées pour les futures nœuds technologiques, d'après les prévisions de l'ITRS.

ment, on peut citer l'OFT (Optical Fourier Transform) permettant une mesure simultanée des intensités diffractées dans tous les angles d'incidence et d'azimut à longueur d'onde fixe [157, 158].

Une fois la signature mesurée, l'étape suivante, souvent plus difficile, consiste à déterminer les dimensions du profil géométrique de la structure par résolution d'un problème inverse. La formulation mathématique du problème est décrite au paragraphe I.3.2 du chapitre I. Sous certaines hypothèses, cette étape consiste en une optimisation paramétrique. Parmi les méthodes employées pour la résolution du problème inverse, on peut citer les méthodes de librairie [148, 159, 160]. Cette méthode nécessite la génération d'une grande quantité de signatures simulées, parmi lesquelles la solution est recherchée. La création de la librairie peut être très longue selon la grandeur de l'espace choisi ainsi que la précision souhaitée. Récemment, certains auteurs ont amélioré cette technique par l'introduction d'algorithmes de régularisation mathématique [161]. Les algorithmes d'optimisation classique (en particulier l'algorithme de Levenberg-Marquardt) sont également largement utilisés dans ce domaine [154, 162, 163, 164].

Les recherches utilisant les réseaux de neurones artificiels pour résoudre le problème inverse en scatterométrie sont relativement récentes. En 1998, Kallioniemi et al. [165] ont employé l'approche neuronale pour la caractérisation des réseaux de diffraction submicroniques (la période étant de l'ordre de 300 nm), de profil géométrique trapézoïdal. La signature optique est formée des intensités en réflexion et en transmission diffractées dans l'ordre 0, mesurées à plusieurs angles d'incidence ($\theta = 40^\circ, 45^\circ, \dots, 65^\circ$) et 2 longueurs d'onde ($\lambda = 442$ nm et 632,8 nm). Ce même groupe de recherche a proposé plus tard une architecture hiérarchique de RN, et a validé la méthode sur un cas plus complexe où le modèle géométrique est décrite par un nombre de paramètres plus grand [166]. Quelques années plus tard, Robert et al. [167, 151] ont validé expérimentalement la méthode avec une signature optique composée d'intensités mesurées à une seule longueur d'onde en fonction de l'angle d'incidence pour différents ordres de diffraction et de polarisations différentes.

La plupart des études citées concernaient essentiellement les réseaux 1D. La caractérisation des réseaux 2D de très faible dimension commence également à émerger depuis peu [168, 169]. Des efforts considérables en scatterométrie ont été réalisés ces 5 dernières années pour la mesure de la rugosité des lignes nuisible pour les composants à très faibles dimensions [170, 171]. La scatterométrie en temps réel est également une voie en cours de développement [161]. Un nouveau mode de mesure est récemment proposé, qui consiste à tremper l'échantillon dans un liquide afin d'améliorer la sensibilité de la mesure du profil géométrique [163].

La première partie de ce chapitre était destinée à la présentation de l'état actuel de l'avancement technologique auprès de l'industrie de la microélectronique au sujet de la réalisation de composants (à finalité optique ou électronique) de plus en plus intégrés. Il est clair que le développement d'une méthode capable de répondre aux besoins des nœuds technologiques prévus pour le futur, est le point clef pour toute avancée technologique.

La partie suivante est consacrée à l'implication des RN pour résoudre le problème inverse dans le processus de caractérisation d'un réseau de diffraction.

III.2 Scatterométrie par réseau de neurones

Nous allons détailler la mise au point d'un PMC destiné à la résolution du problème inverse en scatterométrie. Il s'agit de la détermination des paramètres géométriques définissant un profil périodique de réseaux de diffraction. Nous démontrerons les performances

de la méthode neuronale sur des réseaux possédant différents types de profils géométriques. Les résultats seront comparés, dans la mesure du possible, avec ceux obtenus par les méthodes microscopiques (AFM et MEB). Puis, l'utilisation d'un PMC pour le contrôle d'homogénéité de l'échantillon sera finalement présentée.

III.2.1 Problème inverse en scatterométrie

Comme il a été dit auparavant, la scatterométrie consiste à déterminer le profil géométrique d'une structure périodique à partir d'une signature optique mesurée directement sur l'échantillon par résolution du problème inverse de la diffraction (figure III.10). Dans le cas général, ce problème est mal posé et un certain nombre d'hypothèses doivent être réalisées pour garantir l'unicité de la solution et diminuer également le nombre de paramètres à déterminer. Ainsi, les valeurs des indices optiques sont généralement supposées connues. La forme du profil voire la période peut également être fixée. La formulation paramétrique de chaque profil permet de se ramener à un problème d'optimisation paramétrique ou de régression.

Lorsque le PMC est utilisé en scatterométrie, son rôle est d'apprendre la relation liant la signature ellipsométrique \vec{y} aux différents paramètres définissant le profil géométrique de la structure étudiée. Ainsi, le vecteur d'entrée au PMC n'est autre que l'ensemble des mesures composant la signature et les sorties sont les paramètres géométriques recherchés définissant le profil géométrique.

Considérons le cas d'un réseau de diffraction de silice de période connue $\Lambda = 565$ nm. Le profil géométrique du réseau est supposé trapézoïdal symétrique défini par les paramètres b_1 , b_2 et h (figure I.14). Les domaines de variation de ces paramètres sont choisis de la façon suivante :

$$\begin{aligned} 0 < b_1 &< 60 \text{ nm} \\ 60 < b_2 &< 160 \text{ nm} \\ 150 < h &< 220 \text{ nm} \end{aligned} \tag{III.2}$$

La signature ellipsométrique utilisée est composée de 46 mesures, qui sont les intensités I_s et I_c mesurées en configuration classique (équation I.36) à angle d'incidence classique ($\varphi = 0^\circ$, $\theta = 60^\circ$, pour les longueurs d'onde comprises entre 350 nm et 800 nm par pas de 20 nm).

La base des données d'entraînement du PMC contient 2000 réseaux simulés dont les paramètres géométriques b_1 , b_2 et h sont tirés aléatoirement dans les domaines de variation définis par III.2. La signature ellipsométrique présentée est calculée (paragraphe I.2.2.2) pour chacun des réseaux simulés, pour former ainsi les couples d'entraînement {paramètres géométriques / signature ellipsométrique}, soit $\{\vec{y}, \vec{a}\}$ d'après les notations utilisées précédemment dans ce manuscrit. L'ensemble des signatures calculées sont simulées avec un bruit gaussien de moyenne nulle et d'écart type $\sigma = 3.10^{-3}$. Le choix de la méthode de l'arrêt prématuré pour l'entraînement (partie II.3.3.3) impose que cet ensemble soit divisé en 3 parties distinctes : 70% pour les données d'apprentissage, 20% pour la validation et 10% pour les données test utilisées en dehors de la procédure d'apprentissage dans le but d'estimer les performances non biaisées du PMC (partie II.3.3.2).

Rappelons que les performances E_{test} évaluées sur l'ensemble test donnent seulement une évaluation moyenne globale et ne reflètent pas l'erreur spécifique à chaque échantillon.

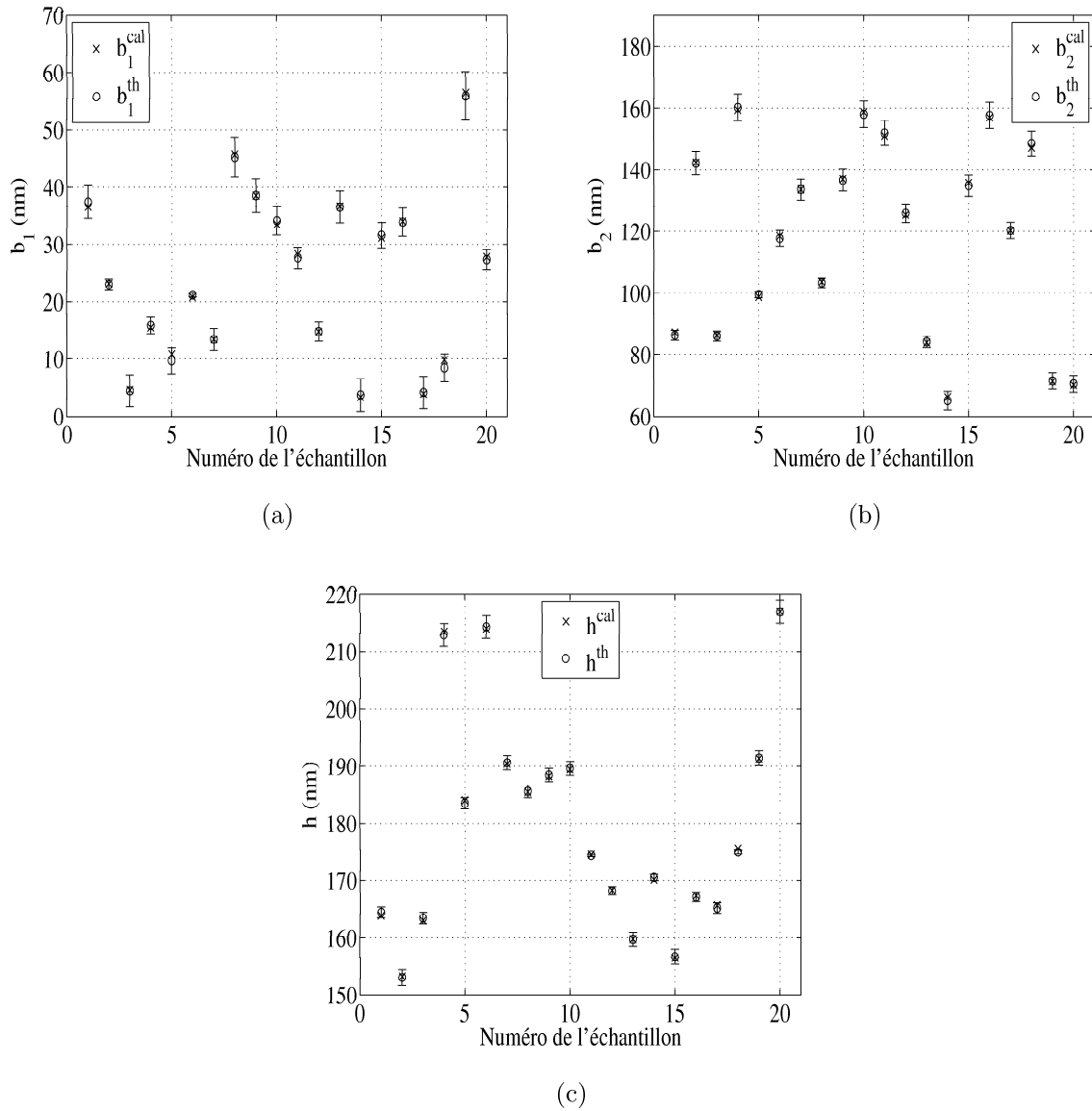


FIG. III.12: Estimation des paramètres géométriques (a) b_1 , (b) b_2 et (c) h par le PMC, ainsi que les barres d'incertitude représentées à 1σ , pour un ensemble de 20 réseaux simulés compris dans le sous domaine test du corpus.

Pour cela, un deuxième PMC est implémenté pour estimer les intervalles de confiance sur les paramètres s donnés par la caractérisation neuronale (paragraphe II.5.1). Ces deux PMC vont donc estimer les paramètres géométriques ainsi que les barres d'erreurs associées, pour chaque nouvelle signature.

La figure III.12 montre les paramètres théoriques connus et les paramètres estimés par le RN pour 20 échantillons simulés tirés de l'ensemble test, ainsi que les barres d'incertitude à 1σ associées à chacun des paramètres (b_1 , b_2 et h). Les barres d'incertitude sont censées recouvrir la valeur vraie dans 68% des cas, ce qui est vérifié ici. On voit apparaître des zones pour chacun des paramètres où les performances du PMC sont différentes. Par exemple, la précision sur le paramètre b_2 est meilleure lorsque sa valeur est comprise entre 80 nm et 120 nm que dans le reste du domaine.

Le PMC est donc capable d'approcher la relation entre la signature ellipsométrique et

les paramètres géométriques correspondant, dans l'espace défini par III.2. Il peut désormais être appliqué expérimentalement.

III.2.2 Etude d'un réseau de silice

Cette étude fait l'objet d'une caractérisation expérimentale d'un réseau de silice de période supposée connue $\Lambda = 565$ nm, d'après la valeur de la période propre au masque. Ce réseau a été fabriqué au sein du Laboratoire Hubert Curien à Saint Etienne dans le cadre du projet européen NEMO (Network of Excellence on Micro-Optics). En premier lieu, nous détaillerons les résultats issus de mesures microscopiques (AFM et MEB). Ensuite, une caractérisation par PMC sera effectuée et les divers résultats seront finalement comparés.

III.2.2.1 Caractérisations microscopiques

Les deux types de caractérisations (AFM et MEB) ont été réalisés par les différents partenaires du projet. Deux caractérisations MEB, notées MEB1 et MEB2, ont été effectuées respectivement au Department of Physics (Université de Joensuu, Finlande) et au Department of Applied Physics and Photonics (Vrije Universiteit Brussel, Bruxelles). Celles de l'AFM, notées AFM1 et AFM2, sont issues respectivement du laboratoire Hubert Curien (Université Jean Monnet, France) et du partenaire Belge déjà cité.

Les résultats correspondants au MEB1 et MEB2 sont donnés par les images de la figure III.13. Rappelons que cette caractérisation a été effectuée en dernier lieu puisqu'elle nécessite la destruction de l'échantillon.

A partir de ces images brutes de l'échantillon données à très basse échelle, l'utilisateur peut juger la forme du profil géométrique du réseau. Cependant, il faut être prudent puisque ce résultat dépend de plusieurs facteurs comme la résolution de l'image, la qualité du clivage et l'alignement du réseau. Il faut noter aussi qu'il est très difficile de réaliser la mesure au même endroit dans le but de confronter plusieurs caractérisations d'un même échantillon.

En supposant la forme du profil trapézoïdale symétrique et à l'aide d'une procédure de traitement d'images propre à chaque partenaire, les paramètres géométriques correspondants sont donnés dans le tableau III.1 pour les deux images MEB.

Les résultats de la caractérisation réalisée par les AFM sont représentés par la figure III.14. La première (pour AFM1) est effectuée sur une surface de $3\mu\text{m} \times 3\mu\text{m}$, la deuxième (pour AFM2) sur une surface de $5\mu\text{m} \times 5\mu\text{m}$. A partir de ces images, les profils bruts sont extraits. Les paramètres géométriques sont ensuite mesurés à l'aide de curseurs en supposant de la même façon que précédemment une forme trapézoïdale symétrique. Les résultats sont regroupés dans le tableau III.1.

Remarque : Il est difficile de chiffrer une barre d'incertitude car les résultats sont fournis tels quels par les partenaires.

Plusieurs enseignements sont à tirer de ces résultats : Tout d'abord, la valeur de la pente b_1 obtenue varie significativement suivant l'une ou l'autre des deux techniques microscopiques. L'AFM mesure une valeur beaucoup plus grande que celle donnée par le MEB. En effet, cela est dû essentiellement à la forme de la pointe de l'AFM qui influe de

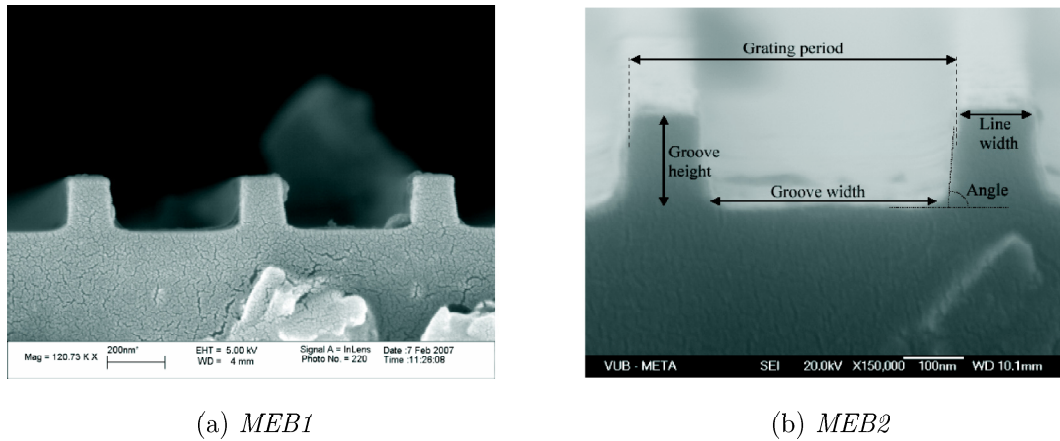


FIG. III.13: Images du réseau en silice provenant de deux microscopes électronique à balayage différents notés (a) MEB1 (Department of Physics, Université de Joensuu, Finlande) et (b) MEB2 (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles).

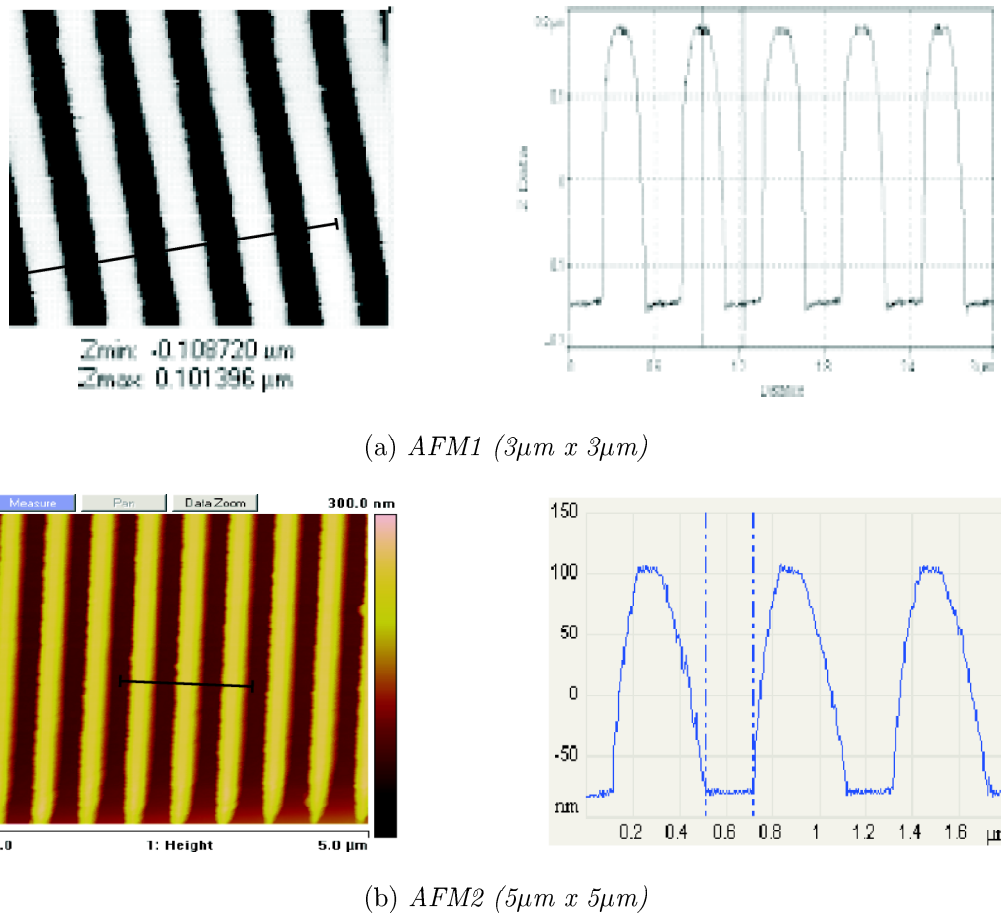


FIG. III.14: Images en 3 dimensions mesurées par deux microscopes à force atomique différents notés (a) AFM1 (Laboratoire Hubert Curien, Saint Etienne) et (b) AFM2 (Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Bruxelles), ainsi que l'extraction sur quelques périodes des profils géométriques à partir de chacune de ces images.

façon nette sur la forme du profil essentiellement au niveau des flancs (voir paragraphe III.1.3.1).

Paramètres géométriques	MEB1	MEB2	AFM1	AFM2
b_1 (nm)	20	25	100	116
b_2 (nm)	111	105	100	98
h (nm)	166	179	189	180
Λ (nm)	543	584	578	609

TAB. III.1: Paramètres géométriques b_1 , b_2 et h et période Λ extraits des images obtenues par MEB et AFM, dans le cas d'un modèle de profil trapézoïdal.

De plus, la différence notable de résultats d'une même méthode entre deux partenaires peut provenir du fait que les mesures ne sont pas réalisées au même endroit de la surface de l'échantillon et éventuellement du traitement des images qui n'est pas identique pour extraire le profil.

Nous remarquons aussi que la valeur de la période diffère non négligemment pour les 4 mesures effectuées, ce qui constitue une réelle surprise. En effet, celle-ci est supposée mesurable de façon très précise.

Après avoir présenté les résultats de ces deux méthodes microscopiques, nous allons passer à la phase de caractérisation par scatterométrie.

III.2.2.2 Caractérisation par réseau de neurones

Cette caractérisation indirecte nécessite en premier lieu la mesure de la signature ellipsométrique. Celle-ci est définie au paragraphe III.2.1. Pour la détermination des paramètres géométriques du réseau de diffraction, la forme trapézoïdale symétrique du profil est supposée. Deux voies sont alors explorées : la première suppose la période fixe, ($\Lambda = 565$ nm) et les paramètres (b_1 , b_2 et h) sont à déterminer, la deuxième cherche à estimer les 4 paramètres (b_1 , b_2 , h et Λ) ensemble.

Période fixe Dans ce cas, nous allons reprendre le PMC présenté dans le paragraphe III.2.1. La signature ellipsométrique mesurée est fournie en entrée du PMC, noté RN_1 qui calcule instantanément les paramètres géométriques b_1 , b_2 et h regroupés dans le tableau III.2.

Période variable Il arrive cependant que la période diffère de façon notable de sa valeur nominale espérée lors des étapes technologiques dû à différentes causes lors du processus de gravure (défaut d'alignement du masque, etc...). Dans notre cas, nous avons remarqué que les caractérisations MEB et AFM concernant la période diffèrent de plusieurs dizaines de nanomètres. Ainsi, il est nécessaire que la méthode de caractérisation utilisée puisse déterminer la période du réseau afin de se confronter aux résultats des méthodes microscopiques.

Deux manières sont possibles pour déterminer la période par voie neuronale : la première consiste à entraîner un PMC, que l'on notera RN_2 , afin d'estimer en sortie les 4 paramètres b_1 , b_2 , h et Λ , comme le schématise la figure III.15(a). La deuxième nécessite l'utilisation de deux PMC découplés, notés RN_{31} et RN_{32} , entraînés indépendamment pour estimer respectivement les 3 paramètres b_1 , b_2 et h ainsi que la seule période Λ , comme le montre le schéma de la figure III.15(b).

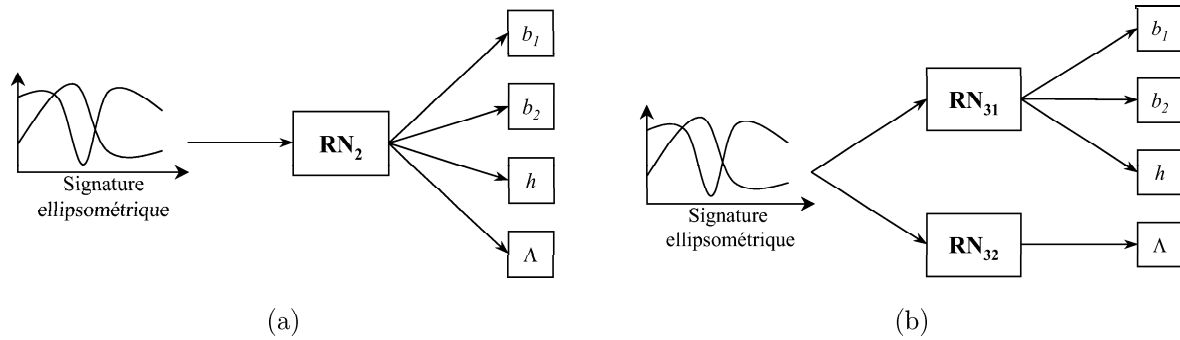


FIG. III.15: Détermination respective des paramètres géométriques du réseau de diffraction (b_1 , b_2 et h) ainsi que sa période Λ par 2 méthodes neuronales (a) utilisant un seul PMC, noté RN_2 (b) utilisant 2 PMC, notés RN_{31} et RN_{32} .

L'entraînement de chaque PMC nécessite $n = 2000$ échantillons simulés dont les paramètres géométriques sont tirés aléatoirement dans les domaines suivants :

$$\begin{aligned} 0 < b_1 &< 60 \text{ nm} \\ 60 < b_2 &< 160 \text{ nm} \\ 150 < h &< 220 \text{ nm} \\ 550 < \Lambda &< 580 \text{ nm} \end{aligned} \quad (\text{III.3})$$

Le nombre N_c de neurones cachés est pris égale à 20 pour les réseaux RN_2 et RN_{31} , et $N_c = 15$ pour RN_{32} plus restreinte. Les résultats donnés par les différents PMC utilisés sont regroupés dans le tableau III.2. Dans chaque cas un second PMC est bien entendu utilisé pour estimer les intervalles de confiance à 1σ associés à chaque paramètre.

Paramètres géométriques	RN_1	RN_2	RN_{31} et RN_{32}
b_1 (nm)	$37,2 \pm 2$	$27,4 \pm 3$	$26,5 \pm 3$
b_2 (nm)	$109,3 \pm 2,8$	$120 \pm 4,2$	$119 \pm 4,5$
h (nm)	$185,5 \pm 1,5$	$178,7 \pm 1,5$	$179 \pm 1,5$
Λ (nm)	x	570 ± 3	$574,6 \pm 4,2$

TAB. III.2: Paramètres géométriques b_1 , b_2 , h et Λ obtenus, selon un profil trapézoïdal symétrique, à partir des 4 PMC notés RN_1 , RN_2 , RN_{31} et RN_{32} .

Notons tout d'abord, que les résultats sont similaires entre RN_2 et RN_3 (RN_{31} et RN_{32}). Le découplage de l'information en sortie de RN n'est pas donc pas forcément nécessaire dans ce cas. Ensuite, il n'est pas surprenant de voir les résultats différer suivant la prise en compte ou non de la période fixe de 565 nm (environ 10 nm pour b_1 et b_2 et 5 nm pour h). En effet, la période influe grandement sur la signature ellipsométrique. Enfin, si on compare les valeurs obtenues à celles du tableau III.1, les résultats sont sensiblement plus proches de ceux du MEB (en particulier *MEB2*) que ceux de l'AFM.

Il est cependant intéressant d'estimer lequel des profils obtenus par les différentes techniques énoncées correspond le plus au "vrai" profil du réseau. Pour cela, nous avons calculé les erreurs quadratiques moyennes sur I_s et I_c entre des mesures réalisées sur l'échantillon et celle calculée à partir des paramètres estimés par chacune des différentes techniques.

Il est important de préciser, qu'afin d'éviter tout biais, ce calcul est effectué dans une configuration expérimentale différente de celle utilisée dans le processus de caractérisation propre par RN. Ces mesures correspondent à I_s et I_c pour $\theta = 55^\circ$, $\varphi = 0^\circ$ et $350 \text{ nm} < \lambda < 800 \text{ nm}$ par pas de 10 nm. Les erreurs ainsi calculées sont tracées sur la figure III.16.

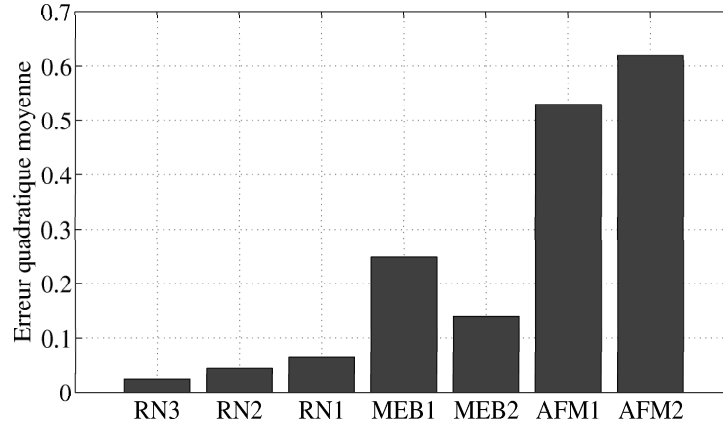


FIG. III.16: Représentation de l'erreur quadratique moyenne calculée sur I_s et I_c à partir de 92 mesures non utilisées pour les profils reconstruits à partir de la méthode neuronale (RN_1 , RN_2 et RN_3), l'AFM (AFM 1 et AFM 2) et le MEB (MEB 1 et MEB 2).

Les erreurs obtenues pour les deux MEB sont strictement inférieures à celles données par les deux AFM. En effet, le profil déterminé par AFM n'est autre que la convolution entre le profil réel du réseau et la forme de la pointe utilisée. Ceci conduit inévitablement à des résultats fortement erronés.

Les résultats obtenus par la méthode neuronale aboutissent à des erreurs relativement faibles comparativement aux deux autres méthodes. Cependant, il ne faut pas tirer de conclusions trop pointues car la validation choisie ici est basée sur le même principe que la méthode utilisée, à savoir la diminution d'une erreur quadratique moyenne. Il se peut également que le profil ne soit pas strictement un trapèze, comme nous l'avons supposé ; ce qui ne peut que nuire aux méthodes microscopiques.

Notons enfin que d'après les valeurs des trois erreurs relativement faibles données par les différents RN, nous pouvons confirmer que la période n'est pas exactement égale à 565 nm comme espéré mais plutôt proche de 572 nm.

D'après notre critère, le profil estimé par RN_2 et RN_3 (RN_{31} et RN_{32}) correspond davantage au profil réel du réseau expérimental. Toutefois, dans le cas du RN_2 , le nombre de poids ajustables dans le PMC augmente et la surface d'erreur risque d'être plus complexe. Il est ainsi conseillé d'utiliser l'architecture de la figure III.15(b), où un PMC est dédié à la seule caractérisation de la période Λ .

Nous venons de présenter la méthode neuronale appliquée en scatterométrie ellipsométrique pour la caractérisation expérimentale des réseaux de diffraction de faible période. Nous avons également fourni une méthode simple permettant de déterminer la période avec une bonne précision (quelques nanomètres). Avant de généraliser la méthode sur des profils géométriques plus ou moins complexes, attardons nous sur quelques considérations concernant le temps nécessaire à un traitement par PMC.

III.2.3 Considérations sur la durée de traitement

Pour une architecture donnée d'un PMC, les performances sont essentiellement régies par deux paramètres : le nombre de neurones N_c dans la couche cachée et le nombre d'exemples n utilisés pour l'entraînement. Bien que la phase d'exploitation par le PMC entraîné soit instantanée, il est intéressant de voir comment influe le choix de ces deux paramètres sur le temps d'apprentissage.

Pour cela, considérons le PMC précédent dédié à la caractérisation de réseaux de diffraction en silice pour une période fixe $\Lambda=565$ nm.

La figure III.17 donne une idée du temps nécessaire pour effectuer un entraînement de ce réseau en fonction du nombre n d'exemples utilisés dans la base d'apprentissage pour différents nombre N_c de neurones cachés. Ce temps correspond au temps CPU de l'ordinateur utilisé (Pentium IV, CPU 2GHz, RAM 1Go).

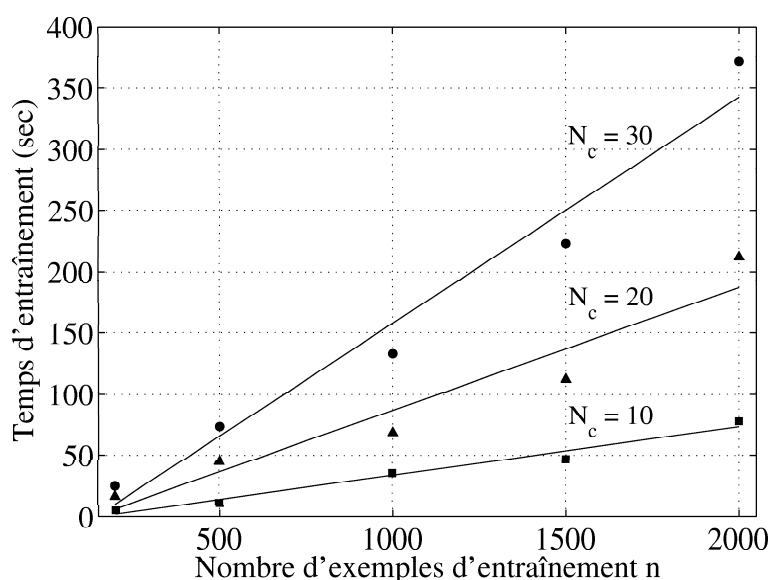


FIG. III.17: Temps d'entraînement du PMC en fonction du nombre n d'exemples d'apprentissage et du nombre N_c de neurones.

Il est évident que le temps d'entraînement augmente lorsque le nombre de neurones augmente, pour un nombre n fixe. En effet, le nombre de poids ajustables dans le réseau augmente, à savoir $46 \times 10 \times 3 = 1380$ pour 10 neurones, $46 \times 20 \times 3 = 2760$ pour 20 neurones et $46 \times 30 \times 3 = 4140$ pour 30 neurones.

D'un autre côté, le temps augmente d'une façon linéaire en fonction du nombre d'exemples. Il faut noter que le nombre des poids ajustables est le même quel que soit le nombre de données d'entraînement. Par contre, si n augmente, cela nécessite plus de calcul d'erreur et de gradient lors de l'entraînement ; ce qui explique l'augmentation du temps en fonction de n .

Ainsi, le paramètre temps n'est pas à négliger pour la phase d'apprentissage. Il est directement lié au nombre de poids ajustables et au nombre de couples utilisés. Toutefois, si la recherche du nombre de poids minimal est également primordiale pour une bonne généralisation, il n'en est pas de même pour le nombre de couples. En effet, plus le nombre n est important meilleur sera la représentation de l'hyperplan engendré par le vecteur d'entrée et plus faible seront les risques d'interpolation erronée. On a donc intérêt à privilégier

un nombre n important tout en gardant des considérations temporelles raisonnables.

III.2.4 Caractérisation de différents profils géométriques

Comme pour toutes les méthodes de caractérisation par résolution du problème inverse, nous avons vu que la méthode neuronale nécessite l'hypothèse a priori d'une forme géométrique de profil. Ainsi, elle peut être facilement appliquée à différents profils lorsque ceux-ci sont préalablement définis. Nous allons à présent montrer la capacité de cette méthode à caractériser des réseaux de diffraction ayant différents profils géométriques.

III.2.4.1 Réseau de profil sinusoïdal

Ce type de profil est souvent obtenu par les techniques de fabrication par interférence. Il est défini par deux paramètres géométriques (figure III.18) : la demi période b de la sinusoïde et la hauteur h . Il faut noter que dans ce cas la période du réseau Λ est différente de celle de la sinusoïde (sauf pour $b = \Lambda/2$).

Remarque : Notons que ce profil supposé peut être plus modulable en ajoutant un troisième paramètre qui est la hauteur de coupe de la sinusoïde.

Le réseau étudié a été fabriqué au sein du Laboratoire Hubert Curien (UMR CNRS 5516, Saint Etienne) à partir d'un masque de résine de période $\Lambda = 2 \mu\text{m}$ déposé sur un substrat de verre BK_7 par photolithographie optique. La forme sinusoïdale est ensuite artificiellement obtenue par chauffage au delà de sa température de transition. Notons que les indices de réfraction de la résine et du substrat BK_7 utilisés sont donnés dans l'annexe A. Le choix d'une période aussi grande est double : d'une part faciliter la caractérisation du réseau par l'AFM en limitant les artefacts dus aux dimensions et d'autre part pour traiter une grande variété de réseaux différents par la méthode neuronale.

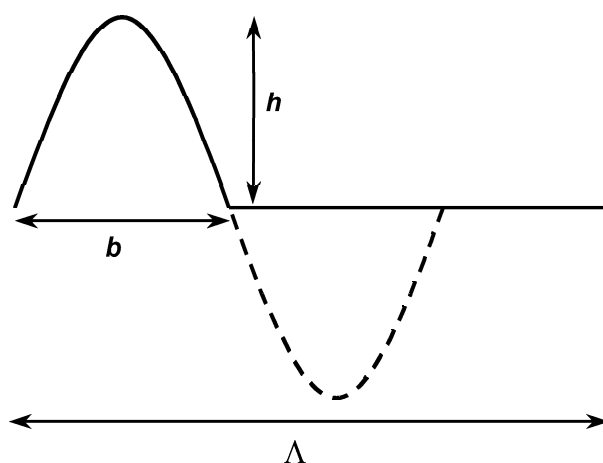


FIG. III.18: Profil sinusoïdal de période Λ défini par deux paramètres géométriques b et h .

Pour la mise en œuvre du PMC dédié à la caractérisation de ce type de profil, définis-

sons le domaine d'entraînement suivant :

$$\begin{aligned} 1300 < b < 1800 \text{ nm} \\ 250 < h < 450 \text{ nm} \end{aligned} \quad (\text{III.4})$$

Le corpus d'entraînement est formé de 1500 couples dont les paramètres b et h sont tirés aléatoirement dans ces domaines. La signature ellipsométrique utilisée correspond aux intensités I_s et I_c mesurées en incidence classique pour $\theta = 45^\circ$, pour une série de longueur d'ondes comprise entre 350 nm et 700 nm par pas de 20 nm. Le PMC contient donc 36 entrées et 2 sorties. Le nombre de neurones cachés est fixé à 20.

Une fois l'entraînement achevé, les performances du réseau sont évaluées par le calcul des erreurs sur 100 nouveaux échantillons test tirés aléatoirement dans le même domaine. Les erreurs E_{test} estimées sont de l'ordre de 1,1 nm pour b et 0,3 nm pour h .

La signature mesurée expérimentalement sur l'échantillon est ensuite introduite en entrée du PMC qui estime alors les paramètres $b = 1631 \pm 3,8$ nm et $h = 325,7 \pm 1,5$ nm.

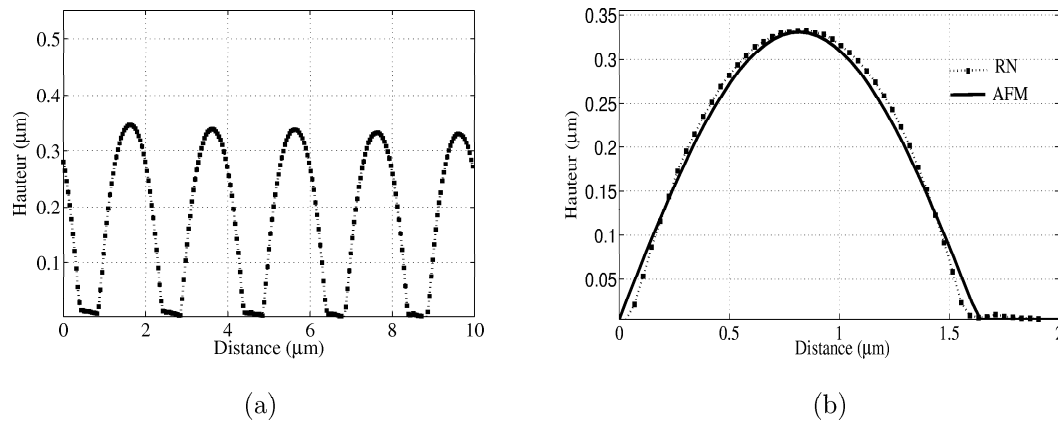


FIG. III.19: (a) Résultat d'une caractérisation AFM du réseau sinusoïdal sur une ligne de 10 μm, ainsi qu'une comparaison (b) sur une période, entre le profil AFM et celui estimé par le PMC.

Afin d'évaluer la qualité du résultat, nous avons comparé le profil obtenu avec celui issue d'une caractérisation par AFM. Le profil du réseau sur une distance de 10 μm est montré sur la figure III.19(a). Une représentation sur une période des deux profils ainsi obtenus est reportée sur la figure III.19(b). En fait, le profil obtenu par AFM n'est pas déformé dans ce cas puisque la période est grande. Le profil est donc correctement modélisé par une fonction sinusoïdale, confirmé par la concordance des résultats.

III.2.4.2 Réseau de profil trapézoïdal symétrique à bords arrondis

En pratique, il est rare d'obtenir des réseaux ayant un profil parfaitement trapézoïdal symétrique. Ainsi, nous avons opté pour un profil géométrique plus complexe, mais qui décrit mieux les échantillons rencontrés en pratique. Ce modèle est basé sur le profil trapézoïdal symétrique auquel sont ajoutés des bords arrondis en haut et en bas des sillons. Il est ainsi défini par 5 paramètres (figure III.20) : les 3 paramètres du trapèze b_1 , b_2 , h agrémentés des rayons de courbure bas R_1 et haut R_2 .

L'échantillon de période $\Lambda = 2 \mu\text{m}$, a été fabriqué dans le même cadre que le réseau sinusoïdal, utilisant la même technique de photolithographie et le même type de matériau. Nous allons effectuer la caractérisation selon deux hypothèses de profils géométriques : trapézoïdal symétrique simple et trapézoïdal symétrique arrondi.

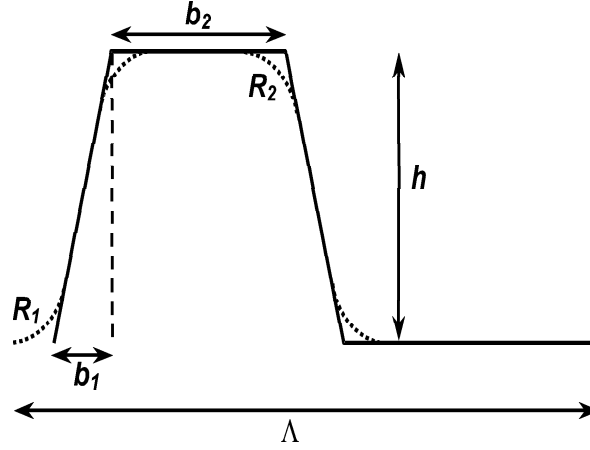


FIG. III.20: Profil trapézoïdal à bords arrondis défini par 5 paramètres : la projection de la pente b_1 , la largeur de la ligne b_2 , la hauteur h et les rayons de courbures bas et haut (respectivement R_1 et R_2).

Nous allons considérer que les paramètres géométriques varient dans les domaines suivants, pour le profil trapézoïdal simple :

$$\begin{aligned} 0 < b_1 < 100 \text{ nm} \\ 700 < b_2 < 900 \text{ nm} \\ 400 < h < 600 \text{ nm} \end{aligned} \quad (\text{III.5})$$

et pour le profil trapézoïdal arrondi :

$$\begin{aligned} 0 < b_1 < 100 \text{ nm} \\ 700 < b_2 < 900 \text{ nm} \\ 400 < h < 600 \text{ nm} \\ 0 < R_1 < 100 \text{ nm} \\ 0 < R_2 < 150 \text{ nm} \end{aligned} \quad (\text{III.6})$$

La signature ellipsométrique utilisée est identique au précédant cas. Le nombre d'échantillons d'entraînement est fixé à $n = 2000$ pour les deux PMC, notés RN_a et RN_b , destinés respectivement à caractériser le réseau selon le profil trapézoïdal simple et le profil arrondi. Le nombre de neurones cachés est pris égal à $N_c = 20$ pour RN_a , et $N_c = 25$ pour RN_b puisque le nombre des paramètres à estimer est plus grand. Les performances E_{test} des deux PMC sont données dans le tableau III.3.

Il est normal de trouver que les erreurs sur les rayons de courbure R_1 et R_2 soient plus grand que celles des autres paramètres. En effet, ces deux paramètres sont généralement qualifiés de second ordre, c'est-à-dire qu'ils influent moins que les autres paramètres sur la signature ellipsométrique.

Ces deux PMC sont désormais appliqués à la caractérisation de l'échantillon décrit

	E_{test}				
	$b_1(\text{nm})$	$b_2(\text{nm})$	$h(\text{nm})$	$R_1(\text{nm})$	$R_2(\text{nm})$
RN_a	2	3,5	1,5	x	x
RN_b	2,8	3,4	1,5	6	9

TAB. III.3: Erreurs quadratiques moyennes calculées pour 100 échantillons test pour les PMC, notés RN_a et RN_b , dédiés à la caractérisation respective de réseaux de profil trapézoïdal symétrique simple (RN_a) et trapézoïdal symétrique arrondi (RN_b).

précédemment. Les paramètres géométriques calculés à partir de chaque PMC sont regroupés dans le tableau III.4, et les profils correspondants sont donnés dans la figure III.21(a).

	$b_1(\text{nm})$	$b_2(\text{nm})$	$h(\text{nm})$	$R_1(\text{nm})$	$R_2(\text{nm})$
RN_a	$51,7 \pm 3,9$	$798,5 \pm 9,9$	$459,8 \pm 3,6$	x	x
RN_b	$48,5 \pm 5,4$	$770,3 \pm 5,3$	$470,3 \pm 2,5$	$70,4 \pm 12$	$33,5 \pm 16$

TAB. III.4: Paramètres géométriques du profil de l'échantillon en résine déposé sur un substrat de BK_7 estimés par la méthode neuronale selon un profil trapézoïdal symétrique et un profil trapézoïdal symétrique arrondi.

Dans ce cas, les profils semblent relativement proches. Cependant, nous avons comparé les intensités calculées à partir des paramètres géométriques du tableau III.4 et celles mesurées dans d'autres configurations expérimentales (différentes valeurs de l'angle d'azimut φ). Les erreurs quadratiques moyennes entre les intensités calculées et celles mesurées sont données sur la figure III.21(b). Dans tous les cas, le profil arrondi est celui qui engendre l'erreur la plus faible quel que soit la valeur de φ . Ceci confirme que le réseau expérimental n'est pas parfaitement trapézoïdal, mais plutôt proche d'un profil à bords faiblement arrondis.

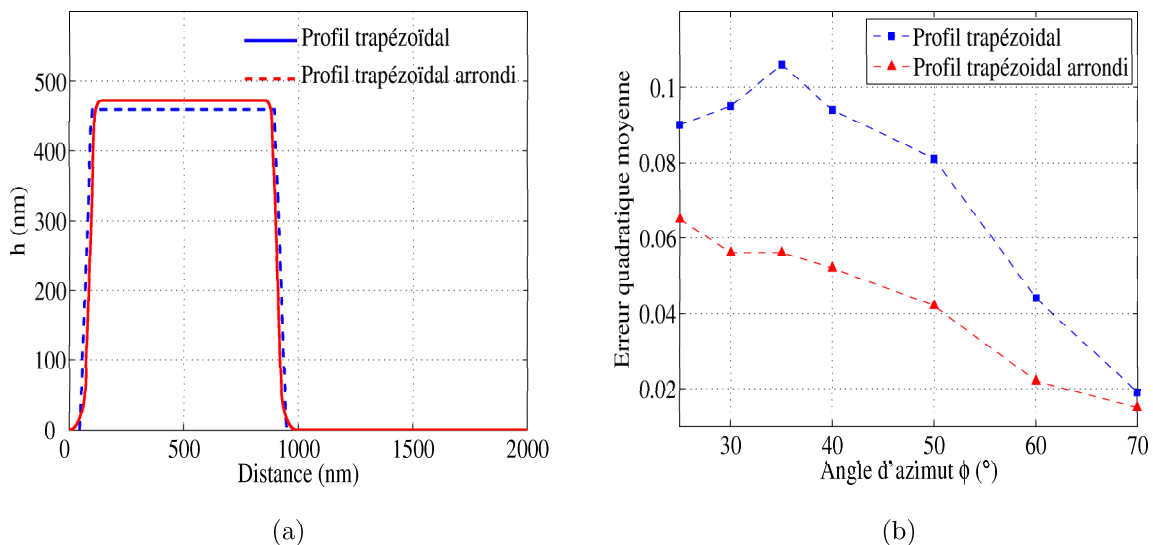


FIG. III.21: (a) Profils géométriques obtenus par RN_a et RN_b , (b) Erreur quadratique moyenne calculée pour différentes configurations autre que celle utilisé pour l'entraînement des RN.

Les efforts consentis ici pour affiner le profil ne s'avèrent pas payants vu les dimensions

des courbures confrontées à la forte valeur de la période. Ce profil sera la plupart du temps considéré comme trapèze symétrique.

Cette recherche fine de la forme du profil implique la détermination de paramètres d'ordre supérieur. Ceci nécessite alors la modification de la signature ellipsométrique par la recherche de mesures porteuses d'information plus riche. Enfin, notons que si la détermination d'un profil précis et détaillé est primordial pour certaines applications (comme amélioration des technologies de fabrication par élimination d'artefacts ainsi visualisés), elle l'est moins pour d'autres qui se contentent généralement de la mesure d'un ou plusieurs paramètres (le plus connu est sans contexte le paramètre de ligne b_2 ou CD).

III.2.4.3 Réseau de profil trapézoïdal symétrique avec une couche résiduelle

Ce profil présente un double intérêt en pratique : pour certaines techniques de gravure, la couche résiduelle peut être considérée comme un défaut de fabrication du masque de résine déposé. En effet, un réseau qui ne débouche pas sur le substrat est considéré comme incompatible pour l'étape de gravure ultérieure. Pour d'autres techniques, comme la nanoimpression par exemple, cette couche résiduelle est laissée volontairement au fond du réseau après pressage du moule dans la résine (paragraphe III.1.2) pour éviter tout contact entre le moule et le substrat. Elle doit être constante tout le long de la surface de pressage afin de garantir une gravure uniforme dans le substrat, d'où la nécessité de la caractériser avec précision.

Les échantillons caractérisés pour ce type de profil sont fabriqués au sein du Laboratoire des Technologies de la Microélectronique LTM (UMR 5129, Grenoble) par la technique de nanoimpression thermique et possèdent différentes périodes. Le moule est réalisé sur un substrat de silicium 8 pouces (200 mm). Les échantillons de résine sont transférés par pressage du moule dans une couche de résine photosensible déposée sur un substrat de silicium de même taille. Les indices de réfraction de la résine et du silicium sont donnés dans l'annexe A. Le profil du réseau formé sur la plaque de Silicium est défini par 4 paramètres géométriques : b_1 , b_2 , h ainsi que l'épaisseur de la couche résiduelle h_r (figure III.22).

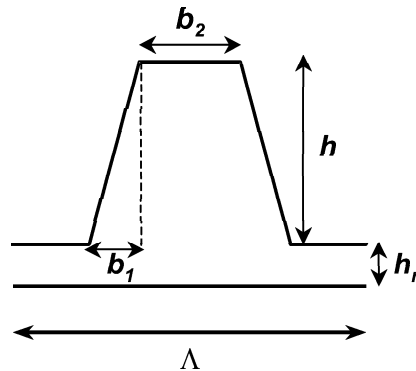


FIG. III.22: Profil trapézoïdal avec une couche résiduelle au fond des sillons, défini par 4 paramètres : b_1 , b_2 , h et l'épaisseur de la couche résiduelle h_r sur le substrat.

Le premier échantillon étudié est noté $E_{h_r}^1$. Sa période est donnée approximativement égale $\Lambda = 665$ nm. Nous avons opté pour l'architecture neuronale présentée précédemment sur la figure III.15(b).

Les images MEB de cet échantillon sont données par la figure III.23, montrant la présence de la couche résiduelle au fond du réseau. Nous pouvons constater également que le réseau est bien trapézoïdal, justifiant le choix du profil défini sur la figure III.22.

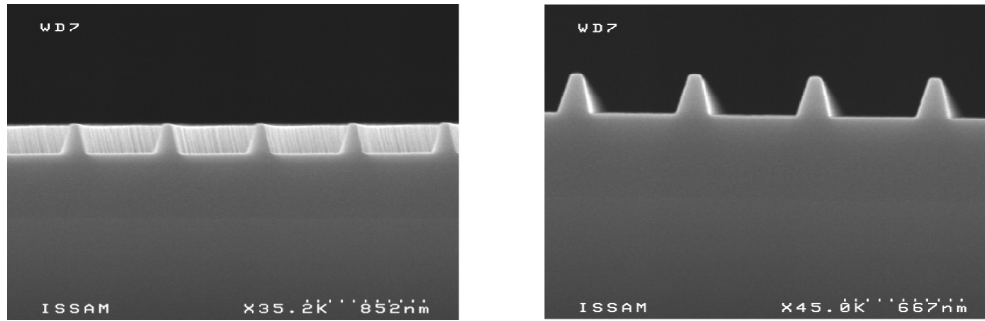


FIG. III.23: Images de profil obtenues par une caractérisation MEB sur le réseau en résine noté E_{hr}^1 .

La signature ellipsométrique est composée de 46 mesures d'intensités I_s et I_c ($\theta = 60^\circ$, $\lambda = 350 : 20 : 800 \text{ nm}$), constituant les entrées au PMC utilisé pour la caractérisation. En suivant le principe de la figure III.15(b), les deux PMC utilisés sont entraînés avec $n = 3000$ échantillons. Les domaines de variation des paramètres sont donnés, pour RN_{31} par :

$$\begin{aligned} 0 < b_1 &< 90 \text{ nm} \\ 20 < b_2 &< 120 \text{ nm} \\ 225 < h &< 275 \text{ nm} \\ 500 < h_r &< 550 \text{ nm} \end{aligned} \quad (\text{III.7})$$

et pour RN_{32} par :

$$660 < \Lambda < 700 \text{ nm} \quad (\text{III.8})$$

Les résultats donnés par les 2 PMC considérés sont regroupés dans le tableau III.5. Rappelons que la détermination des valeurs de ces paramètres à partir des images MEB est délicate. Ainsi, nous n'en feront pas état dans ce qui suit.

RN_{31}				RN_{32}
$b_1(\text{nm})$	$b_2(\text{nm})$	$h(\text{nm})$	$h_r(\text{nm})$	$\Lambda(\text{nm})$
$69,3 \pm 8,3$	$51,4 \pm 1,2$	$252,6 \pm 4,1$	$520,5 \pm 1,5$	$675,1 \pm 4,6$

TAB. III.5: Paramètres géométriques et période du profil de l'échantillon E_{hr}^1 estimé par la méthode neuronale selon un profil trapézoïdal avec une couche résiduelle.

Dans la partie précédente, la validité des résultats a été démontrée par l'usage du problème direct. Une autre façon de vérifier la véracité des résultats consiste à réaliser plusieurs caractérisations distinctes à partir de signatures ellipsométriques disjointes et d'étudier la stabilité des résultats obtenus.

Ceci nécessite évidemment la mise au point de 3 PMC distincts, entraînés par exemple sur la même signature que précédemment mais prise pour différents angles d'azimut (à savoir ici 20° , 40° et 60°). Les paramètres géométriques accompagnés de leurs barres d'incertitude obtenus à partir des 3 PMC ainsi créés sont représentés sur la figure III.24 et

comparés à ceux obtenus pour $\varphi = 0^\circ$. Les valeurs de chacun des paramètres obtenues par chaque PMC sont presque identiques et incluses dans les barres d'incertitude correspondantes. Nous pouvons conclure que le profil trapézoïdal supposé est juste, ainsi que les paramètres trouvés pour le réseau expérimental. Les valeurs moyennes de chaque paramètre sont : $b_1 = 69,1$ nm, $b_2 = 51,9$ nm, $h = 252,4$ nm, $h_r = 520,5$ nm et $\Lambda = 675$ nm.

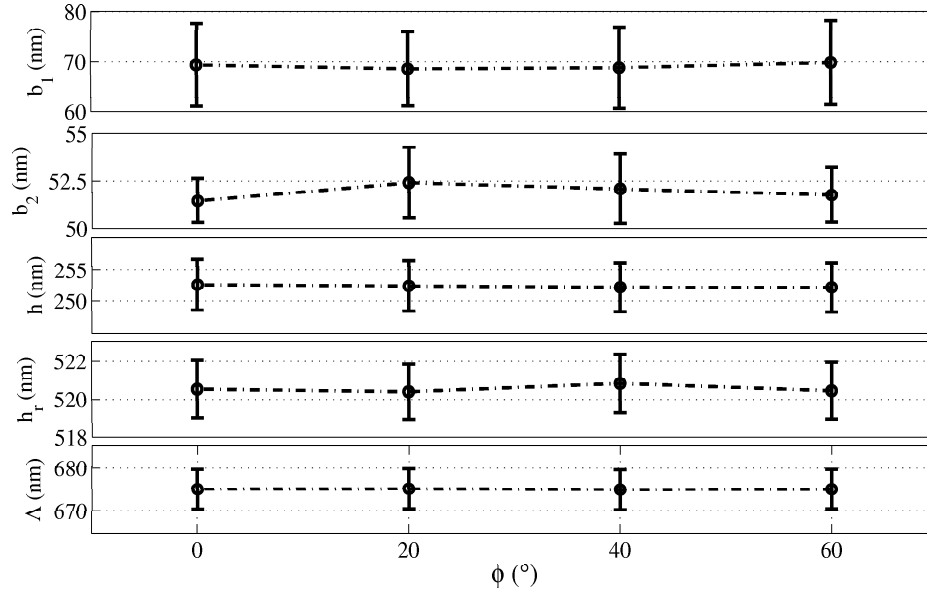


FIG. III.24: Paramètres géométriques calculés pour différentes valeurs de l'angle d'azimut φ par 4 caractérisations neuronales distinctes.

Dans le même cadre d'étude, nous allons caractériser 3 réseaux expérimentaux, notés $E_{h_r}^2$, $E_{h_r}^3$ et $E_{h_r}^4$ fabriqués dans des conditions bien déterminées. Le but est de montrer que l'épaisseur de la couche résiduelle h_r peut être déterminée avec précision, comme le souhaitent les utilisateurs de la technique de nanoimpression. La hauteur initiale de la couche de résine déposée sur le substrat de silicium est égale $h_i = 215$ nm. Les propriétés géométriques souhaitées des réseaux sont données dans le tableau III.6.

Echantillon	Période (nm)	$R_{l/e}$	h_r attendue (nm)
$E_{h_r}^2$	1500	600/900	119
$E_{h_r}^3$	2000	1100/900	83
$E_{h_r}^4$	4000	1100/2900	149

TAB. III.6: Descriptif des propriétés attendues des réseaux fabriqués par nanoimpression sur une couche de résine photosensible d'épaisseur $h_i = 215$ nm déposée sur un substrat de silicium.

Notons qu'il est possible de déterminer théoriquement, selon la loi de conservation du volume de la matière déplacée, l'épaisseur de la couche résiduelle qui se forme après pressage. Pour cela, l'épaisseur initiale de la couche de résine déposée sur la plaque à presser et les dimensions du moule utilisé doivent être connues (dans notre cas $h_M = 240$ nm). La figure III.25 montre le principe du pressage et le déplacement de la matière pour remplir les cavités dans le motif.

Lorsque le processus se déroule bien et les cavités sont bien remplies, l'épaisseur de la couche résiduelle peut être calculée à l'aide de la formule suivante :

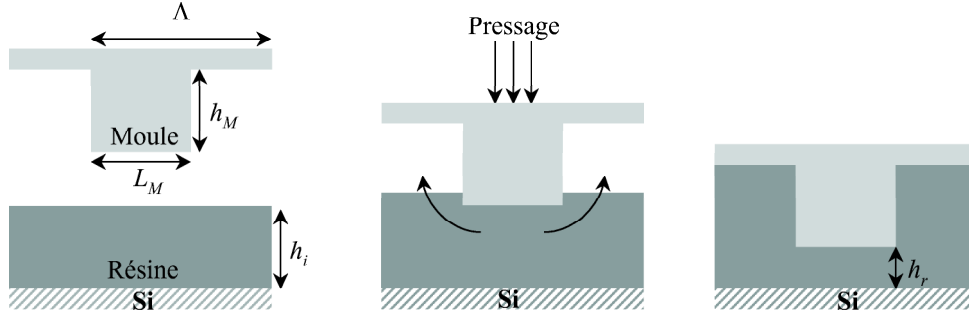


FIG. III.25: Déplacement de la matière lors du pressage dans un processus nanoimpression.

$$h_i \Lambda = h_r \Lambda + (\Lambda - L_M) h_M \Rightarrow h_r = h_i - \frac{(\Lambda - L_M) h_M}{\Lambda} \quad (\text{III.9})$$

avec :

- h_i : épaisseur initiale de la couche de résine déposée sur le substrat avant pressage,
- L_M : largeur de la ligne du motif dans le moule,
- h_M : hauteur du motif dans le moule,
- Λ : période du moule.

Nous allons à présent caractériser ces 3 réseaux par PMC en supposant cette fois-ci la période connue. La signature ellipsométrique est la même que celle utilisée pour l'échantillon $E_{h_r}^1$. Les 3 PMC sont entraînés avec $N_c = 20$ et $n = 3000$ chacun, dans leurs domaines de validité respectifs, notés $D_{h_r}^2$, $D_{h_r}^3$ et $D_{h_r}^4$, donnés dans le tableau III.7. Les résultats de la caractérisation par PMC sont donnés dans le tableau III.8 ainsi que la valeur théorique de h_r obtenue à partir de la loi III.9.

Domaine de validité	Paramètres géométriques			
	b_1 (nm)	b_2 (nm)	h (nm)	h_r (nm)
$D_{h_r}^2$	0 - 80	525 - 650	200 - 270	100 - 180
$D_{h_r}^3$	0 - 80	1025 - 1150	200 - 270	60 - 140
$D_{h_r}^4$	0 - 80	1025 - 1150	200 - 270	120 - 200

TAB. III.7: Domaines de variation des paramètres géométriques b_1 , b_2 , h et h_r , notés $D_{h_r}^2$, $D_{h_r}^3$, $D_{h_r}^4$ utilisés pour construire le corpus d'apprentissage de chaque PMC.

	Paramètres géométriques calculés				h_r théorique (nm)
	b_1 (nm)	b_2 (nm)	h (nm)	h_r (nm)	
$RN_{h_r}^2$	52,3±4,6	552,8± 3,4	236,1± 1,8	118,2±1	119
$RN_{h_r}^3$	47,7±6,3	1073,2±11,4	238,9±1,9	80,3±1,2	83
$RN_{h_r}^4$	23,4±2,8	1068,1±2,4	234,9±1,3	147±1,1	149

TAB. III.8: Paramètres géométriques b_1 , b_2 , h et h_r des échantillons $E_{h_r}^1$, $E_{h_r}^2$ et $E_{h_r}^3$ résultants de la caractérisation neuronale. Comparaison avec la valeur théorique de h_r obtenue par la loi de conservation du volume de la matière déplacée.

Notons que les valeurs des rapports $R_{l/e}$ (tableau III.6) souhaitées ne sont pas forcément obtenues après pressage. Nous constatons également que l'épaisseur h des réseaux

réalisés dans la couche de résine est proche de celle visée, à savoir du moule h_M . De plus, les épaisseurs de la couche résiduelle h_r pour les différents réseaux concordent parfaitement bien avec les valeurs théoriques que nous devons obtenir. Il s'agit dans ce cas là d'une confirmation du bon déroulement de l'étape de nanoimpression.

III.2.4.4 Réseau de profil trapézoïdal asymétrique

En pratique, il est assez rare d'obtenir des défauts conduisant à des profils asymétriques. En effet, les techniques photolithographiques sont assez bien maîtrisées et peuvent facilement éviter des imperfections de ce genre. Cependant, de tels réseaux peuvent être obtenus volontairement par interférence en utilisant la partie non linéaire de la courbe de caractérisation de la résine [172]. L'étude qui va suivre reste dans le cadre théorique devant l'absence de réseau expérimental asymétrique à notre disposition. Le but étant de comprendre comment se comporte le PMC pour ce type de profil, et surtout s'il peut être utilisé pour tester l'asymétrie de certains profils.

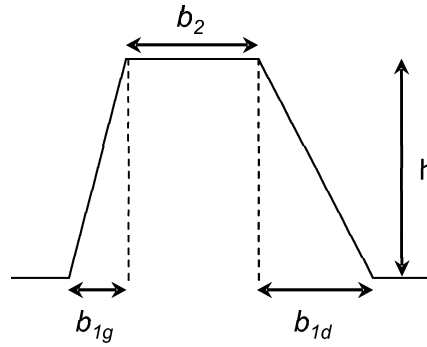


FIG. III.26: Profil trapézoïdal asymétrique défini par 4 paramètres géométriques : les projections des pentes gauches et droites, notées respectivement b_{1g} et b_{1d} , la largeur de ligne b_2 et la hauteur h .

Le profil trapézoïdal asymétrique est défini par 4 paramètres géométriques (figure III.26) : les projections des pentes gauche et droite, notées respectivement b_{1g} et b_{1d} , la largeur de la ligne b_2 et la hauteur h . Nous avons choisi de simuler un réseau en résine de période $\Lambda = 565$ nm déposé sur un substrat de verre BK7. Les indices de réfraction utilisés pour la résine et pour le substrat sont donnés dans l'annexe A.

La signature ellipsométrique, notée C_1 , est composée des intensités I_s et I_c pour $\theta = 60^\circ$ et λ comprise entre 350 nm et 800 nm par pas de 20 nm. Le PMC, noté PMC_1 , contenant $N_c = 20$ neurones cachés est entraîné avec un ensemble de $n = 2000$ échantillons simulés tel que les paramètres géométriques varient selon les domaines suivants :

$$\begin{aligned} 0 < b_{1g} &< 80 \text{ nm} \\ 0 < b_{1d} &< 80 \text{ nm} \\ 90 < b_2 &< 150 \text{ nm} \\ 160 < h &< 200 \text{ nm} \end{aligned} \tag{III.10}$$

Les performances du PMC, données sous forme de l'erreur E_{test} , sont regroupés dans le tableau III.9.

D'après les valeurs élevées des erreurs correspondantes à b_{1g} et b_{1d} , nous pouvons confirmer que le PMC n'est pas capable de déterminer ces deux paramètres avec préci-

	$b_{1g}(\text{nm})$	$b_{1d}(\text{nm})$	$b_2(\text{nm})$	$h(\text{nm})$
E_{test}	17,2	16,8	1,4	0,7

TAB. III.9: Erreurs quadratiques moyennes E_{test} calculées pour 100 échantillons test pour le PMC dédié à la caractérisation de réseau en silice de période $\Lambda = 565 \text{ nm}$ et de profil trapézoïdal asymétrique.

sion, contrairement à b_2 et h . En effet, cela est dû au fait que les réseaux de diffraction asymétriques possèdent un comportement particulier : en diffraction classique, la signature ellipsométrique d'un réseau asymétrique est la même lorsque les projections des pentes droites et gauches b_{1d} et b_{1g} se permutent. De plus, cette signature est presque confondue avec celle donnée par un modèle symétrique tel que la somme des projections des pentes droite et gauche est la même.

Pour illustrer ce phénomène, nous allons simuler la signature ellipsométrique de trois réseaux A_1 , A_2 et A_3 de profils représentés sur la figure III.27(a). Les trois possèdent la même hauteur $h = 200 \text{ nm}$ et la même largeur de ligne $b_2 = 300 \text{ nm}$. La somme des projections de la pente droite et gauche est prise égale à 100 nm , tel que $b_1 = 50 \text{ nm}$ pour le réseau trapézoïdal symétrique A_1 , $b_{1g} = 10 \text{ nm}$ et $b_{1d} = 90 \text{ nm}$ pour le réseau trapézoïdal asymétrique A_2 , $b_{1g} = 90 \text{ nm}$ et $b_{1d} = 10 \text{ nm}$ pour le réseau A_3 . Les signatures ellipsométriques C_1 de ces 3 réseaux particuliers sont tracées sur la figure III.27(b).

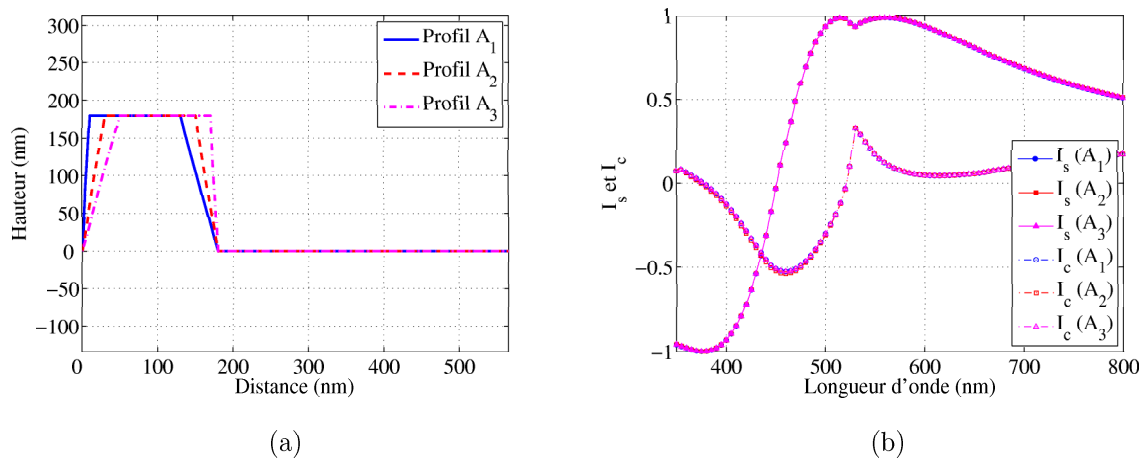


FIG. III.27: (a) Représentation de 3 profils géométriques asymétriques particuliers : A_1 ($b_{1g} = 10 \text{ nm}$, $b_{1d} = 50 \text{ nm}$), A_2 ($b_{1g} = 30 \text{ nm}$, $b_{1d} = 30 \text{ nm}$) et A_3 ($b_{1g} = 50 \text{ nm}$, $b_{1d} = 10 \text{ nm}$), la hauteur h et la largeur de la ligne b_2 sont égales respectivement à 180 nm et 120 nm , la période étant $\Lambda = 565 \text{ nm}$; (b) les signatures ellipsométriques I_s et I_c correspondantes.

Signature ellipsométrique	$\lambda(\text{nm})$	$\theta(^{\circ})$	$\varphi(^{\circ})$	Ordre diffracté
C_1	350 :20 :800	60	0	0
C_2	350 :20 :800	60	0	-1
C_3	350 :20 :800	60	45	0
C_4	400	60	0 :5 :90	0

TAB. III.10: Différentes signatures ellipsométriques choisies en fonction de la longueur d'onde λ , l'angle d'incidence θ , l'angle d'azimut φ et l'ordre de diffraction.

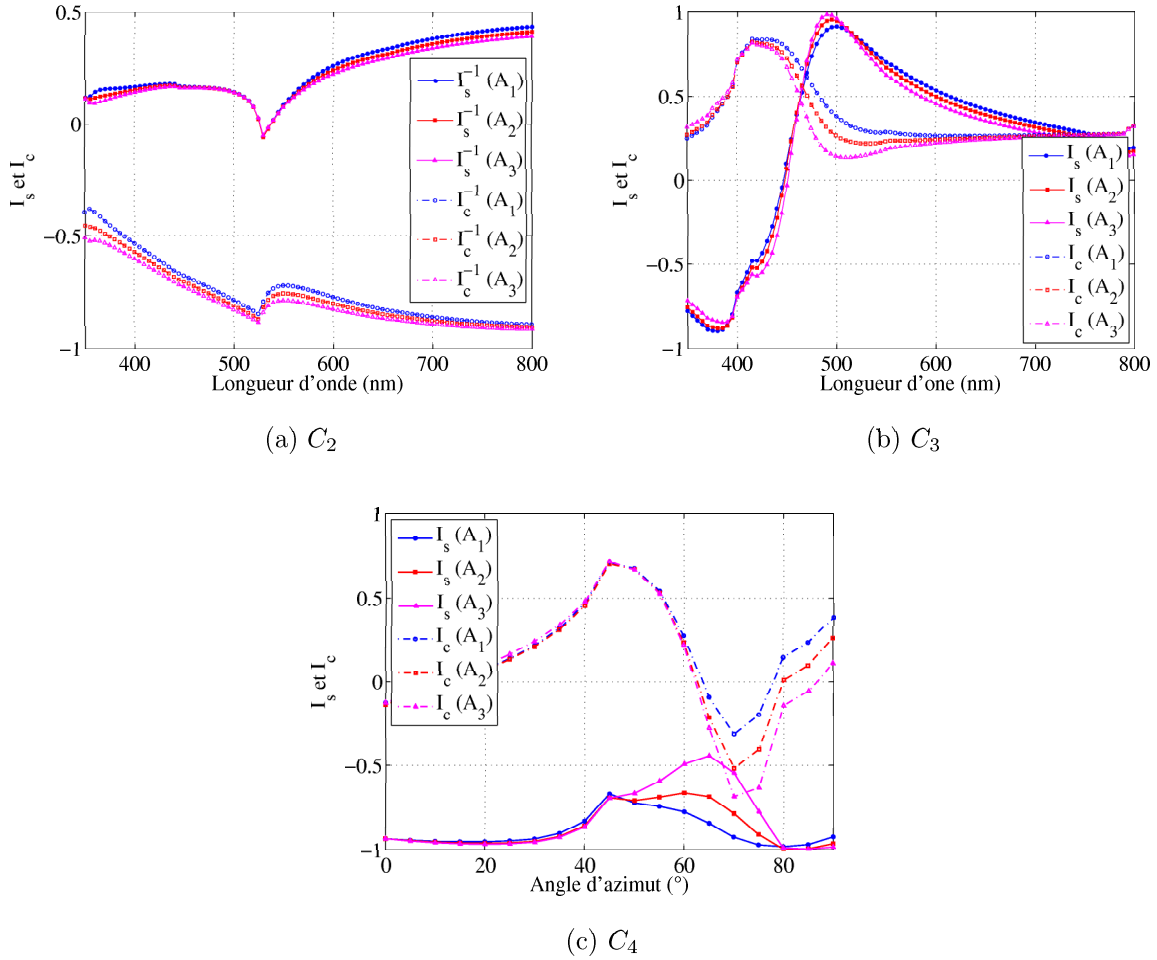


FIG. III.28: Signatures ellipsométriques simulées pour les 3 réseaux (A_1 , A_2 et A_3) de la figure III.27 selon les configurations (C_2 , C_3 et C_4) définies dans le tableau III.10.

D'après ces courbes, nous remarquons qu'en utilisant une signature ellipsométrique mesurée en incidence classique dans l'ordre 0, comme cela a été principalement le cas dans les études précédentes, il n'est pas possible de différencier les 3 profils A_1 , A_2 et A_3 . Dans ce cas, il est difficile de résoudre le problème inverse car la solution n'est pas unique. Le problème est ainsi mal posé.

Dans le but de poursuivre cette étude, différentes signatures ellipsométriques sont définies dans le tableau III.10 pour différentes valeurs de θ , λ et φ . Notons également que la signature C_2 nécessite la mesure d'un ordre différent de la réflexion spéculaire (les problèmes liés à la mesure de cet ordre ne sont pas pris en compte dans cette étude).

La figure III.28 représente les signatures C_2 , C_3 et C_4 simulées respectivement pour les 3 réseaux de profil A_1 , A_2 et A_3 . On voit clairement que ces profils peuvent être décorrélés en utilisant une autre signature ellipsométrique que C_1 notamment, celle mesurée dans un ordre supérieur -1 (C_2), ou celle utilisant une configuration conique (C_3 ou C_4). Il est clair que la configuration C_4 se dégage des 2 autres en particulier dans la zone d'azimut $\varphi = 60^\circ - 80^\circ$. Toutefois, cette constatation reste locale pour un échantillon donné et ne peut pas être généralisé trop rapidement. Il est cependant évident que la signature ellipsométrique est plus sensible à l'asymétrie du profil en configuration conique [156].

Afin de vérifier cette constatation, 3 PMC sont entraînés en utilisant pour chacun

d'eux une des signatures ellipsométriques C_2 , C_3 et C_4 . Le nombre d'échantillons est égal à $n = 2000$ avec $N_c = 20$ neurones cachés pour chaque PMC. Les paramètres géométriques sont susceptibles de varier dans les mêmes domaines de variation utilisés précédemment (III.10). Les performances des différents réseaux entraînés sont comparés dans le tableau III.11.

E_{test}	PMC_2	PMC_3	PMC_4
b_{1g}	1,9	2,9	2
b_2	1,5	1,6	0,8
b_{1d}	2,2	1,7	1,3
h	0,7	0,7	0,5

TAB. III.11: Erreurs quadratiques moyennes E_{test} calculées sur un ensemble de 200 échantillons test pour les 3 PMC utilisant les signatures C_2 , C_3 et C_4 définies dans le tableau III.10.

Les performances évaluées sur chaque PMC sont nettement meilleures que celles du tableau III.9 pour le PMC_1 . Ainsi, le PMC permet d'identifier l'asymétrie des réseaux de diffraction lorsque la signature ellipsométrique utilisée n'est pas en configuration classique. On remarque également de meilleurs résultats dans le cas de l'utilisation de la signature C_4 de longueur d'onde fixe mais d'azimut variable.

Nous avons montré jusqu'à présent la capacité du PMC à caractériser des réseaux de diffraction ayant différents profils géométriques, à partir d'une signature ellipsométrique mesurée. Les réseaux étudiés concernaient plusieurs types de matériaux et de périodes différentes.

La section suivante est consacrée au contrôle de l'homogénéité des échantillons et l'avantage offert par le PMC pour réaliser cette tâche.

III.2.5 Contrôle de l'homogénéité de la structure

Nous avons vu que le traitement des données par un RN entraîné est instantané. Cette caractéristique est mise à profit pour contrôler l'homogénéité d'un réseau de diffraction de période $\Lambda = 400$ nm issu d'un procédé de nanoimpression.

Les dimensions du réseau de diffraction, noté E_{hr}^5 , sont de $5,75$ mm \times $5,75$ mm. Le profil supposé est de forme trapézoïdale avec une couche résiduelle (figure III.22). Les caractérisations ont été effectuées sur des points espacés de 400 μ m de part et d'autre du centre du réseau. Le PMC utilisé est similaire à celui présenté au paragraphe

La figure III.29 représente les variations par rapport au centre des 4 paramètres géométriques b_1 , b_2 , h et h_r .

D'après ces résultats, on peut constater que l'échantillon étudié est quasiment homogène sur une zone qui s'étend sur 1600 μ m \times $2 = 3200$ μ m, malgré une variation décelable qui ne dépasse pas 5 nm sur l'ensemble des paramètres. Par contre, un comportement particulier est observé lorsque l'on s'approche des bords du réseau. La hauteur de la couche résiduelle est supérieure à sa valeur au centre, et le réseau est moins profond sur les bords. En effet, dans certaines conditions de nanoimpression le moule peut subir une déformation [173], comme le montre la figure III.30. Dans la zone du milieu le processus se passe bien et toutes les cavités du moule sont remplies. Par contre, aux bords ce n'est pas le cas et les motifs du moule ne sont pas totalement enfoncés dans la résine. Ainsi, la hauteur des

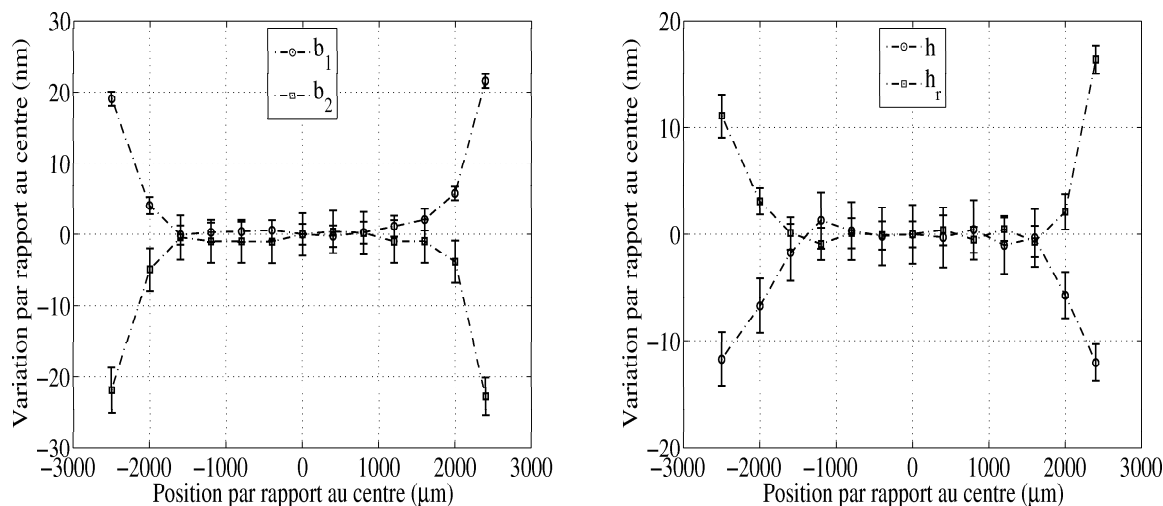


FIG. III.29: Cartographie des paramètres géométriques b_1 , b_2 , h et h_r du réseau $E_{h_r}^5$.

motifs gravés à ces endroits est plus faible qu'au centre alors que l'épaisseur de la couche résiduelle est plus importante.

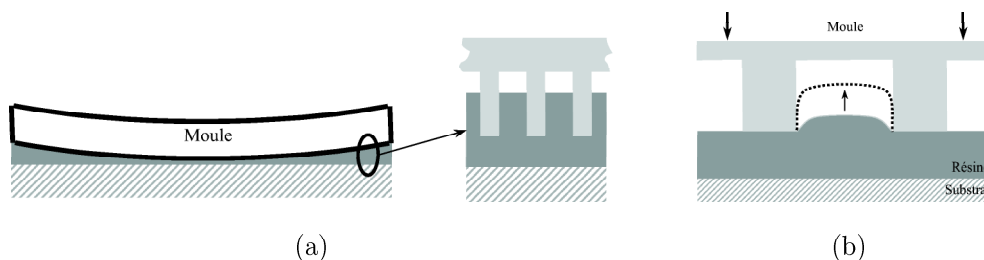


FIG. III.30: Schéma illustrant (a) une déformation du moule dans un processus de nanoimpression et (b) la forme de la résine dans les cavités du moule au fur et à mesure du pressage.

De même, le comportement similaire de b_1 et b_2 concordent avec la pratique puisque la résine ne remplit pas toute la cavité du moule. En fait, le profil de la résine pressée peut être assimilé à un trapèze dont la pente b_1 diminue (tandis que b_2 augmente) au fur et à mesure que la résine remplit les cavités, comme il est illustré dans le schéma de la figure III.30(b).

Le PMC s'avère une méthode puissante comparée aux techniques classiques d'optimisation pour le contrôle de l'homogénéité d'un réseau donné. Le traitement instantané offert par ce dernier est un point fort pour le contrôle en ligne d'un procédé de fabrication sur de larges surfaces. Rappelons que la durée totale d'une caractérisation par un PMC préalablement entraîné est strictement réduit au temps nécessaire à la mesure de la signature ellipsométrique en chaque point.

Nous venons de présenter l'utilisation des RN dans la scatterométrie. Une première étude sur la caractérisation expérimentale d'un réseau de diffraction en silice a été réalisée et comparée avec les techniques microscopiques classiques. Les deux techniques présentées, à savoir l'AFM et le MEB, offrent une caractérisation locale qui peut couvrir quelques périodes. L'extraction du profil est complexe et quelque fois entachée d'erreur. Toutefois,

ces méthodes sont largement utilisées pour des tâches bien précises ; Leur avantage principal est l'absence d'hypothèse sur la forme géométrique. Nous avons également montré la flexibilité des RN qui peuvent facilement être adaptés pour la caractérisation des profils géométriques plus ou moins complexes. Finalement, le traitement instantané offert par les PMC permet un contrôle rapide de l'homogénéité des échantillons.

Les RN sont des outils complexes dont les paramètres intrinsèques sont directement liés à ses performances globales. La partie suivante est consacrée à l'influence de tous ces paramètres directs ou indirects qui influent sur la qualité de la caractérisation optique neuronale.

III.3 Etude paramétrique

Le résultat donné par un PMC est fonction de ses paramètres internes, comme le nombre d'exemples d'entraînement n , le nombre de neurones dans la couche cachée N_c , le domaine de validité des paramètres, etc. . . . Plusieurs hypothèses de départ propres à la structure à caractériser (comme le profil, la signature ellipsométrique) sont également influentes sur les performances du RN.

Les études présentées dans la présente partie sont regroupées en deux sous-parties : la première traite de l'influence des paramètres qualifiés d'internes sur les performances du PMC, alors que la seconde est dédiée plus particulièrement à l'influence du choix de certaines des hypothèses de départ.

III.3.1 Paramètres intrinsèques au réseau de neurones

III.3.1.1 Influence du nombre de données d'entraînement avec le nombre de neurones cachés

Au cours des différentes caractérisations réalisées jusqu'ici par RN, le nombre fixe N_c de neurones dans la couche cachée ainsi que le nombre d'exemples n d'entraînement ont été systématiquement clairement affichés. Il est intéressant d'examiner l'influence de ces deux paramètres sur les performances du PMC lorsque celui-ci est entraîné dans un domaine fixe.

Pour cela, nous allons reprendre le PMC présenté au paragraphe III.2.1 destiné à la caractérisation des paramètres géométriques d'un réseau de diffraction de profil trapézoïdal simple. Ainsi, le nombre des entrées et des sorties est donc respectivement fixé à 46 et 3.

Les erreurs quadratiques E_{test}^{200} calculées sur la base de 200 échantillons test, sont tracées sur la figure III.31 en fonction du nombre n d'exemples d'entraînement pour différents nombres N_c de neurones cachés.

D'après ces courbes, nous constatons que plus le nombre n augmente, plus l'erreur E_{test}^{200} diminue jusqu'à une valeur limite qui dépend du nombre de neurones N_c . En même temps, lorsque le nombre N_c optimal est atteint, l'augmentation de ce dernier n'influe plus sur les résultats. Dans ce cas, le degré de la complexité de la fonction générée par le PMC est suffisant pour approcher efficacement la fonction de régression recherchée dans le domaine fixé. Il reste à s'assurer que le nombre d'exemples n est suffisant pour bien représenter l'espace à modéliser.

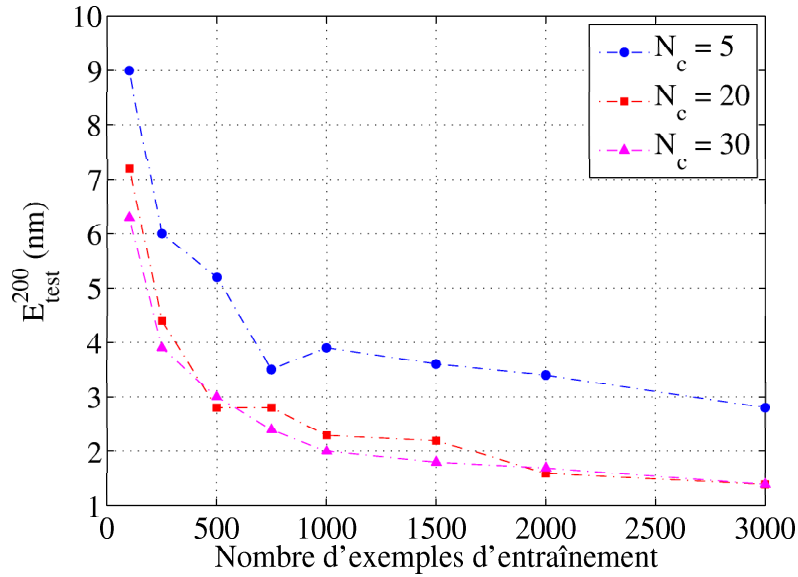


FIG. III.31: Variation des performances du PMC en fonction du nombre n d'exemples d'entraînement pour différents nombres N_c de neurones cachés.

Ainsi, il est inutile de choisir un nombre de neurones trop élevé et accroître ainsi la complexité du réseau. En effet, cela induirait une augmentation du temps d'entraînement (voir figure III.17) du PMC à cause du nombre de poids qui devient important et surtout une augmentation du risque de convergence prématurée.

III.3.1.2 Domaine de validité du réseau de neurones

Il est important de rappeler que le PMC est destiné à fonctionner dans un domaine de l'espace des paramètres géométriques bien défini et fixé par l'utilisateur lors de la phase d'apprentissage. Cette partie est consacrée au comportement du PMC à l'extérieur du domaine. La question sous jacente qui est soulevée ici est la suivante : que se passe-t-il lorsque la signature optique que va traiter le PMC ne correspond pas à un réseau inclus dans la base d'apprentissage.

L'étude est menée en considérant le même PMC du paragraphe précédent, en effectuant la démarche suivante. La base d'apprentissage est définie par l'équation III.2. Les signatures ellipsométriques d'un ensemble d'échantillons de profil trapézoïdal à $b_1 = 30$ nm, $h = 180$ nm et $20 \text{ nm} < b_2 < 200$ nm (par pas de 2 nm) sont simulés puis, introduites en entrée du PMC entraîné. La figure III.32(a) représente la variation relative des 3 paramètres géométriques calculés par le PMC par rapport à leurs valeurs initiales, en fonction de la valeur de b_2 .

La même étude est réalisée cette fois-ci lorsque $100 < h < 270$ nm (par pas de 2 nm), sachant que $b_1 = 30$ nm et $b_2 = 120$ nm sont fixés.

Nous remarquons tout d'abord que le PMC réagit correctement lorsque les paramètres sont tous contenus dans leurs domaines d'entraînement. Par contre, lorsqu'un des paramètres n'appartient pas au domaine pré-cité, les performances du PMC se dégradent et la valeur des paramètres alors calculés est la plupart du temps erronée. Toutefois, il faut signaler que le résultat donné par le PMC reste également correct aux bords des domaines de variation des paramètres.

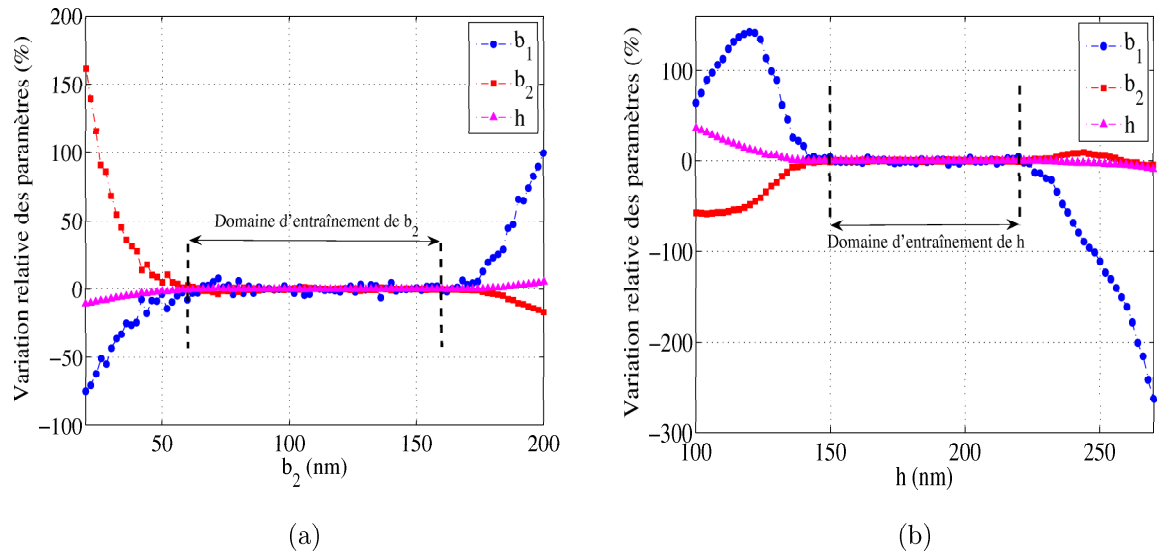


FIG. III.32: Erreurs relatives estimées sur les paramètres géométriques b_1 , b_2 et h pour un échantillon simulé en fonction (a) du paramètre b_2 lorsque $b_1 = 30$ nm et $h = 180$ nm, et (b) du paramètre h lorsque $b_1 = 30$ nm et $b_2 = 120$ nm.

Ainsi, si le PMC calcule en sortie une valeur d'un (ou plusieurs) paramètre(s) géométrique(s) non comprise dans le domaine d'entraînement de ce dernier, il est préférable de modifier le domaine et recommencer la procédure d'apprentissage. Il est donc nécessaire de s'assurer que la solution recherchée se trouve dans l'espace des paramètres spécifié par l'utilisateur.

En conclusion, le réseau de neurones est un système complexe dont l'utilisateur doit bien maîtriser son fonctionnement. Les paramètres internes sont à choisir judicieusement selon la nature et la complexité du problème traité afin que le réseau puisse fournir des résultats fiables et corrects. Dans les paragraphes qui suivent, nous allons étudier l'influence de certains paramètres externes expérimentaux ainsi que ceux liés au choix du modèle géométrique de la structure à caractériser.

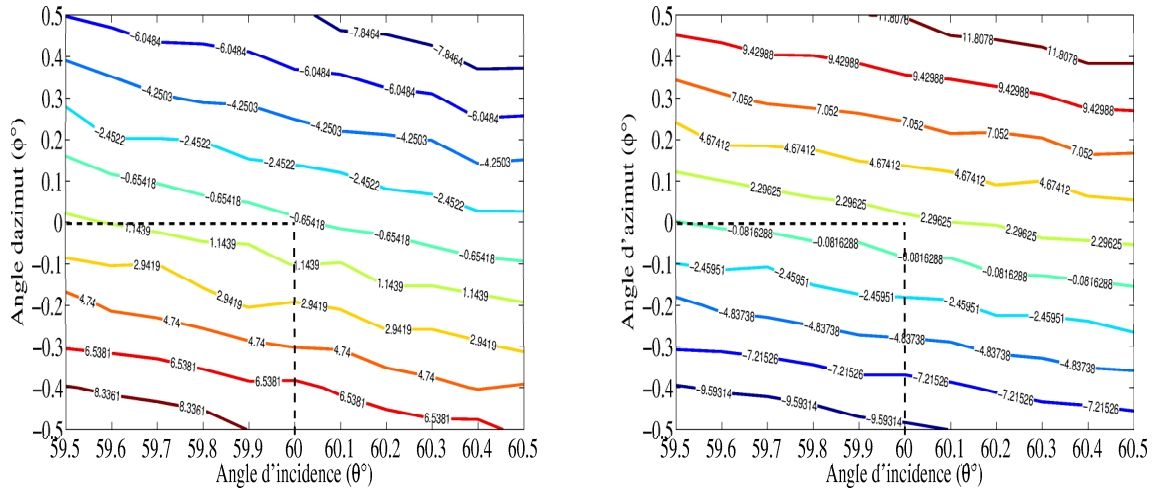
III.3.2 Paramètres externes au réseau de neurones

III.3.2.1 Influence des angles d'incidence θ et d'azimut φ

En pratique, la position des différents éléments constituant le dispositif expérimental de mesure est généralement connue avec une certaine précision. Rappelons que dans notre cas, le réseau de neurones est entraîné avec des signatures optiques calculées pour un angle d'incidence θ et d'azimut φ dont les valeurs sont connues et fixées à l'avance. Dans ce paragraphe, nous allons étudier l'influence de ces deux paramètres lorsque leurs valeurs expérimentales, pour un échantillon donné, diffèrent légèrement de celles prises théoriquement pour entraîner le PMC.

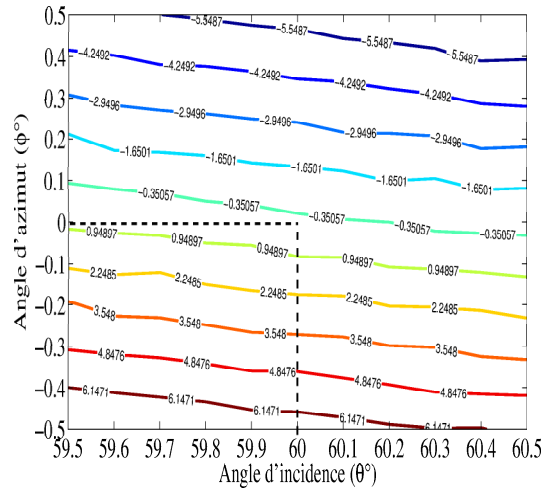
Le même PMC du paragraphe précédent est utilisé, sachant qu'il a été entraîné avec des signatures ellipsométriques simulées pour un angle d'incidence $\theta = 60^\circ$ et d'azimut $\varphi = 0^\circ$. Considérons un réseau de diffraction de paramètres géométriques $b_1 = 30$ nm, $b_2 = 120$ nm et $h = 180$ nm, compris dans les domaines d'entraînement du PMC. Les

signatures ellipsométriques correspondantes sont calculées pour différentes de θ ($59,5^\circ < \theta < 60,5^\circ$ par pas de $0,01^\circ$) et de φ ($-0,5^\circ < \varphi < 0,5^\circ$ par pas de $0,01^\circ$), et sont ensuite introduites au PMC.



(a) Ecart de b_1 en nm par rapport à sa valeur attendue ($b_1 = 30 \text{ nm}$)

(b) Ecart de b_2 en nm par rapport à sa valeur attendue ($b_2 = 120 \text{ nm}$)



(c) Ecart de h en nm par rapport à sa valeur attendue ($h = 180 \text{ nm}$)

FIG. III.33: Variations relatives des paramètres géométriques (a) b_1 (b) b_2 et (c) h calculés par le PMC en fonction d'une erreur systématique sur les angles d'incidence θ et d'azimut φ . Les valeurs nominales de b_1 , b_2 et h étant respectivement 30nm , 120nm et 180nm .

La figure III.33 trace les écarts entre les valeurs attendues de b_1 , b_2 et h , à savoir respectivement 30 nm , 120 nm et 180 nm . Lorsque l'angle d'incidence est légèrement biaisé à angle φ fixe, on remarque que les écarts entre les valeurs calculées des paramètres et leurs valeurs initiales ne sont pas notables. Par contre, une même variation sur l'angle φ à θ fixe influe considérablement sur le résultat. A titre d'exemple, si l'angle φ est connu à $0,5^\circ$ près de sa valeur réelle, on risque de commettre une erreur qui va de 6 nm à 9 nm sur les paramètres géométriques calculés. D'où la nécessité de connaître cet angle avec précision lors de la mesure de la signature optique.

III.3.2.2 Influence de la période

En scatterométrie, la période est souvent traitée comme une constante et supposée connue, fixée pendant le calcul. En général, les processus lithographiques sont suffisamment sophistiqués aujourd'hui pour éliminer la possibilité d'avoir une période incontrôlable. Malgré cela, des erreurs de différents types (liées à la fabrication du masque ou au processus de gravure) peuvent être rencontrées, induisant ainsi un écart de la période obtenue par rapport à la valeur souhaitée.

Il paraît intéressant d'examiner comment réagir un RN entraîné avec un modèle géométrique de période fixe lorsque la signature ellipsométrique introduite en entrée correspond à un réseau de diffraction ayant le même modèle mais une période légèrement différente de celle supposée.

Pour cela, un PMC est entraîné pour estimer en sortie les paramètres géométriques d'un réseau de diffraction en silice de profil trapézoïdal symétrique, de période fixe $\Lambda = 565$ nm. Nous avons considéré 3 réseaux de diffraction ayant pour paramètres $b_1 = 25$ nm, $h = 180$ nm et $b_2 = 326,5$ nm, 270 nm et 175,8 nm de façon à obtenir 3 rapport $R_{l/e}$ différent (équation I.62 page 53), respectivement de 1.5, 1 et 0.5. Le PMC est entraîné avec $N_c = 20$ neurones dans la couche cachée et $n = 3000$ échantillons pris dans les domaines de variation des paramètres suivants :

$$\begin{aligned} 0 < b_1 < 60 \text{ nm} \\ 100 < b_2 < 400 \text{ nm} \\ 150 < h < 220 \text{ nm} \end{aligned} \quad (\text{III.11})$$

Les signatures ellipsométriques pour chacun des 3 réseaux sont simulées pour différentes valeurs de période Λ compris entre 550 nm et 580 nm par pas de 1 nm. Les erreurs relatives des sorties calculées par le PMC, à partir de ces signatures simulées en entrée sont représentées dans la figure III.34 en fonction de la période réelle.

L'hypothèse d'une période fausse n'est pas anodine lors de la détermination des paramètres géométriques d'un réseau de diffraction. L'erreur commise sur ces paramètres est presque indépendante du rapport $R_{l/e}$ du réseau. Elle est moins remarquable sur la hauteur h que sur les deux autres paramètres b_1 et b_2 . A titre d'exemple, un écart de 15 nm sur la période (variation relative de $\pm 2,65\%$) produit une erreur de l'ordre de 5 nm sur la hauteur. Le même écart peut produire une erreur d'environ 25 nm sur le paramètre b_2 . Ainsi, pour une caractérisation précise des paramètres géométriques, la période doit être donnée avec précision. Dans notre cas, ceci peut être achevé à l'aide de la méthode proposée précédemment (paragraphe III.2.2.2).

III.3.2.3 Influence du choix du modèle géométrique

L'hypothèse la plus importante dans la méthode de caractérisation exposée ici est sans aucun doute la forme géométrique de la structure à caractériser. Il est ainsi nécessaire de comprendre le processus de fabrication qui conditionne fortement la forme géométrique obtenue pour approcher au mieux le profil réel.

Le but de cette partie est d'étudier théoriquement l'influence du choix d'un modèle sur le fonctionnement d'un RN.

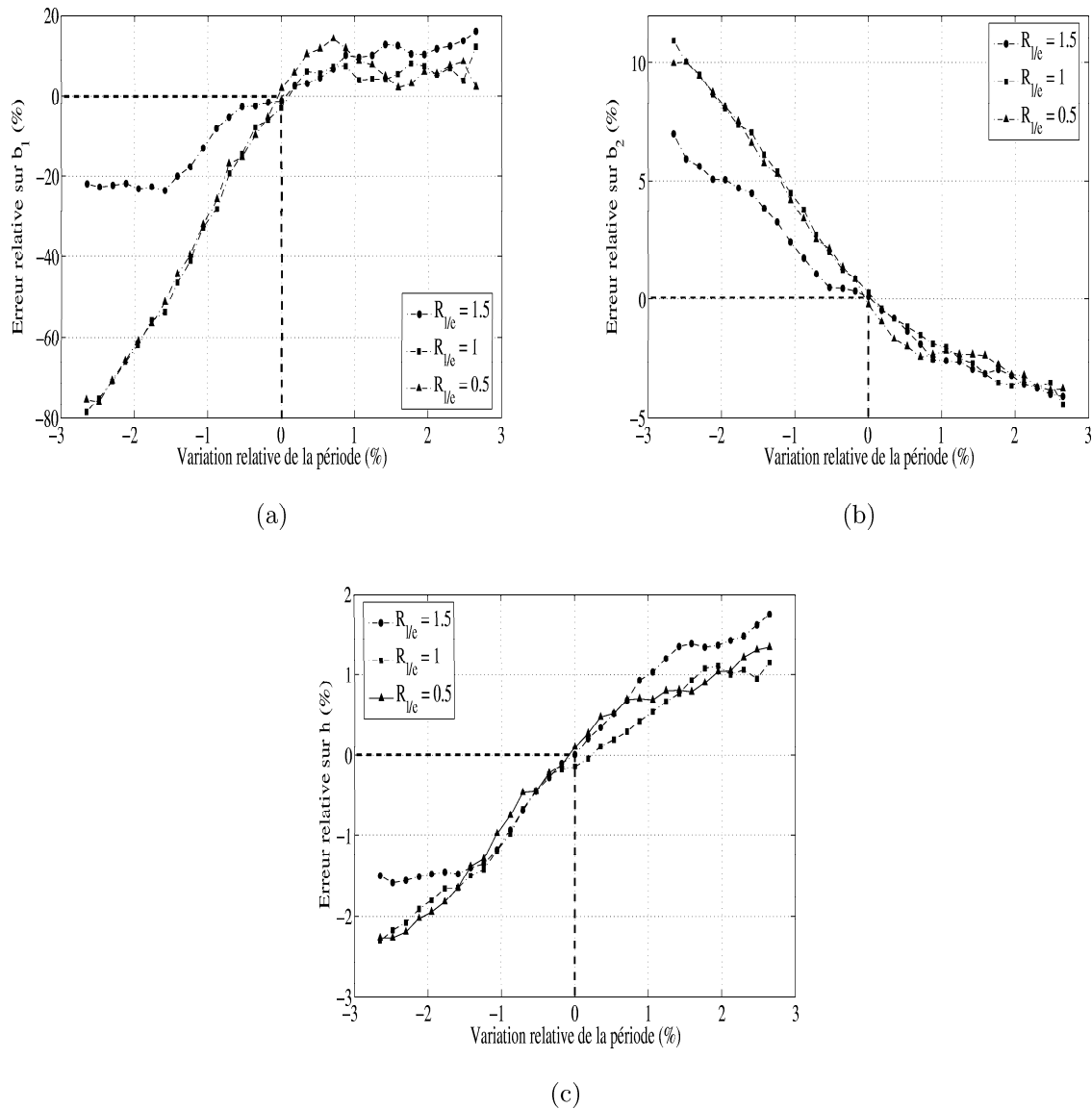


FIG. III.34: Erreur relative des paramètres géométriques (a) b_1 (b) b_2 et (c) h calculés par le PMC en fonction de la variation relative de la période pour différentes valeurs du rapport ligne/espace ($R_{l/e}$). Les valeurs nominales de b_1 et h étant respectivement de 25 nm et 180 nm.

Pour mener cette étude, nous allons suivre les étapes définies sur le schéma de la figure III.35 : Un PMC est d'abord entraîné pour estimer les l paramètres géométriques ($p_1^{geo}, \dots, p_l^{geo}$) d'un profil supposé noté P_2 . Ensuite, nous allons examiner la variation de ces paramètres lorsque nous introduisons en entrée du PMC une signature ellipsométrique correspondante à un autre profil noté P_1 défini par d'autres paramètres.

La signature ellipsométrique utilisée pour cette étude dans les paragraphes qui suivent est composée des 46 mesures définies dans la partie III.2.1. La période est considérée connue.

P_1 : Trapèze symétrique / P_2 : Rectangulaire - Le premier cas étudié est celui où l'on suppose que la structure à caractériser possède un profil rectangulaire, défini par 2 paramètres géométrique b_L et h_L (figure III.36), et gravé sur un substrat de silice. La

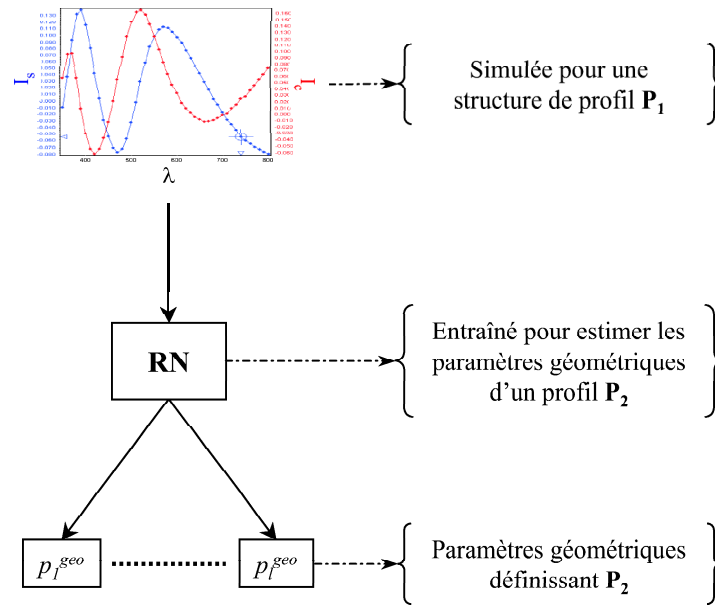


FIG. III.35: Schéma de principe de l'étude concernant l'influence du profil géométrique sur la réponse du PMC.

période étant fixé à $\Lambda = 565$ nm.

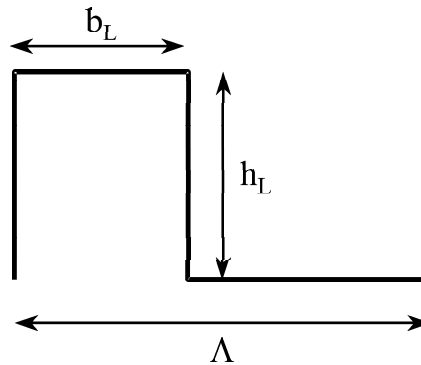


FIG. III.36: Profil rectangulaire défini par deux paramètres géométriques, la largeur de la ligne b_L et la hauteur h_L .

Le PMC utilisé est donc entraîné pour estimer les paramètres b_L et h_L selon le modèle rectangulaire, équivalent à P_2 dans la figure III.35. Les domaines de variation des paramètres sont choisis de la façon suivante :

$$\begin{aligned} 60 < b_L < 160 \text{ nm} \\ 150 < h_L < 220 \text{ nm} \end{aligned} \quad (\text{III.12})$$

Par ailleurs, considérons que la signature ellipsométrique est mesurée en réalité sur un réseau de profil trapézoïdal symétrique, équivalent à P_1 , défini par les 3 paramètres b_1 , b_2 et h (figure I.14 page 52). Ainsi, la présence de la pente b_1 est ignorée par la modélisation neuronale.

Afin de mettre en évidence le comportement du PMC, nous allons considérer une série de réseaux de diffraction simulés dont les paramètres géométriques correspondant au profil P_1 sont $2 \text{ nm} < b_1 < 60 \text{ nm}$, $b_2 = 120 \text{ nm}$ et $h = 180 \text{ nm}$. Les signatures ellipsométriques

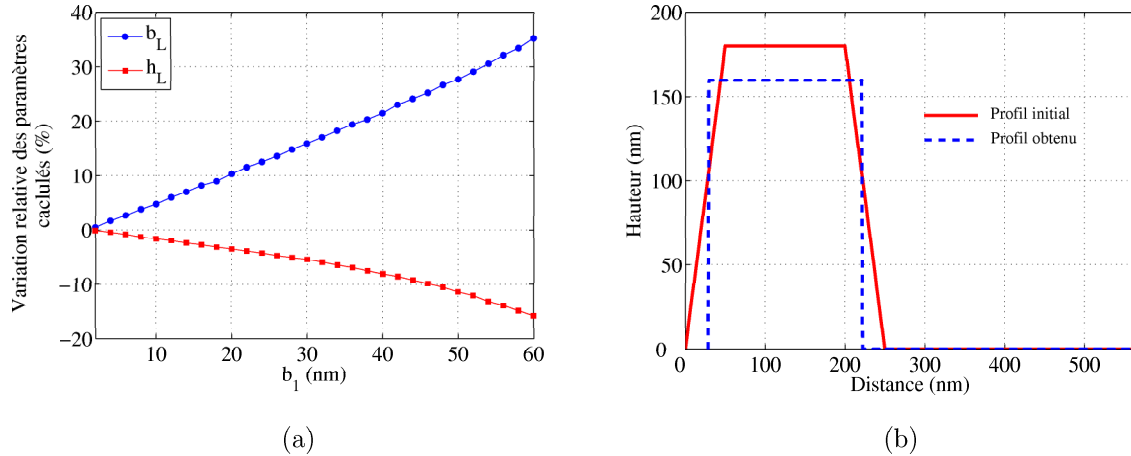


FIG. III.37: (a) Variation relative des paramètres géométriques b_L et h_L par rapport à leurs valeurs nominales, à savoir respectivement 120 nm et 180 nm, en fonction de b_1 . (b) Profil rectangulaire obtenu par le PMC ($b_L = 192.2$ nm et $h_L = 159.6$ nm) à partir d'une signature correspondante à un profil trapézoïdal ($b_1 = 50$ nm, $b_2 = 120$ nm et $h = 180$ nm). La période étant fixé à $\Lambda = 565$ nm.

simulées sont alors introduites dans le PMC préalablement entraîné sur P_2 . La figure III.37(a) représente la variation relative des paramètres (b_L et h_L) calculés par le PMC, par rapport à leurs valeurs exactes en fonction de b_1 .

Lorsque la pente b_1 est ignorée dans la phase de caractérisation, les deux autres paramètres b_L et h_L calculés par le PMC se trouvent modifiés et diffèrent de leurs valeurs exactes. En effet, la largeur de la ligne b_2 tend à croître et la hauteur h diminue de plus en plus (de façon moindre) lorsque la valeur de b_1 augmente et l'erreur sur le profil devient importante. La figure III.37(b) montre un cas particulier d'un profil rectangulaire obtenu par PMC ($b_L = 192.2$ nm et $h_L = 159.6$ nm) pour un profil trapézoïdal initial ($b_1 = 50$ nm, $b_2 = 120$ nm et $h = 180$ nm). En effet, le PMC recherche dans sa base d'apprentissage la structure rectangulaire dont la signature optique correspond le plus à celle fournie en entrée. C'est une optimisation globale des paramètres. Ainsi, le paramètre supposé décrire la même quantité physique dans les 2 profils (à savoir la hauteur h) se trouve également entaché d'une erreur non négligeable.

Remarque : Dans le cas inverse (P_1 correspondant à un profil rectangulaire et P_2 à un profil trapézoïdal), cela ne cause aucun problème puisque le profil P_1 dans ce cas est inclus dans P_2 (pour $b_1 = 0$).

P_1 : Sinus / P_2 : Rectangulaire Le profil du réseau de diffraction à caractériser est à présent sinusoïdal défini par les 2 paramètres géométriques b et h (figure III.18 page 149). Par contre, le PMC est entraîné sur le même profil rectangulaire que précédemment (b_L et h_L).

Afin d'étudier la réponse du PMC, nous avons introduit deux séries de signatures ellipsométriques en faisant varier l'un ou l'autre des paramètres définissant le profil sinusoïdal : la première correspond aux réseaux simulés tel que $b = 120$ nm et 160 nm $< h < 200$ nm (par pas de 2 nm), la deuxième est telle que $b = 100 : 2 : 140$ nm et $h = 180$ nm.

Les variations relatives des paramètres géométriques b et h estimés à l'aide du PMC

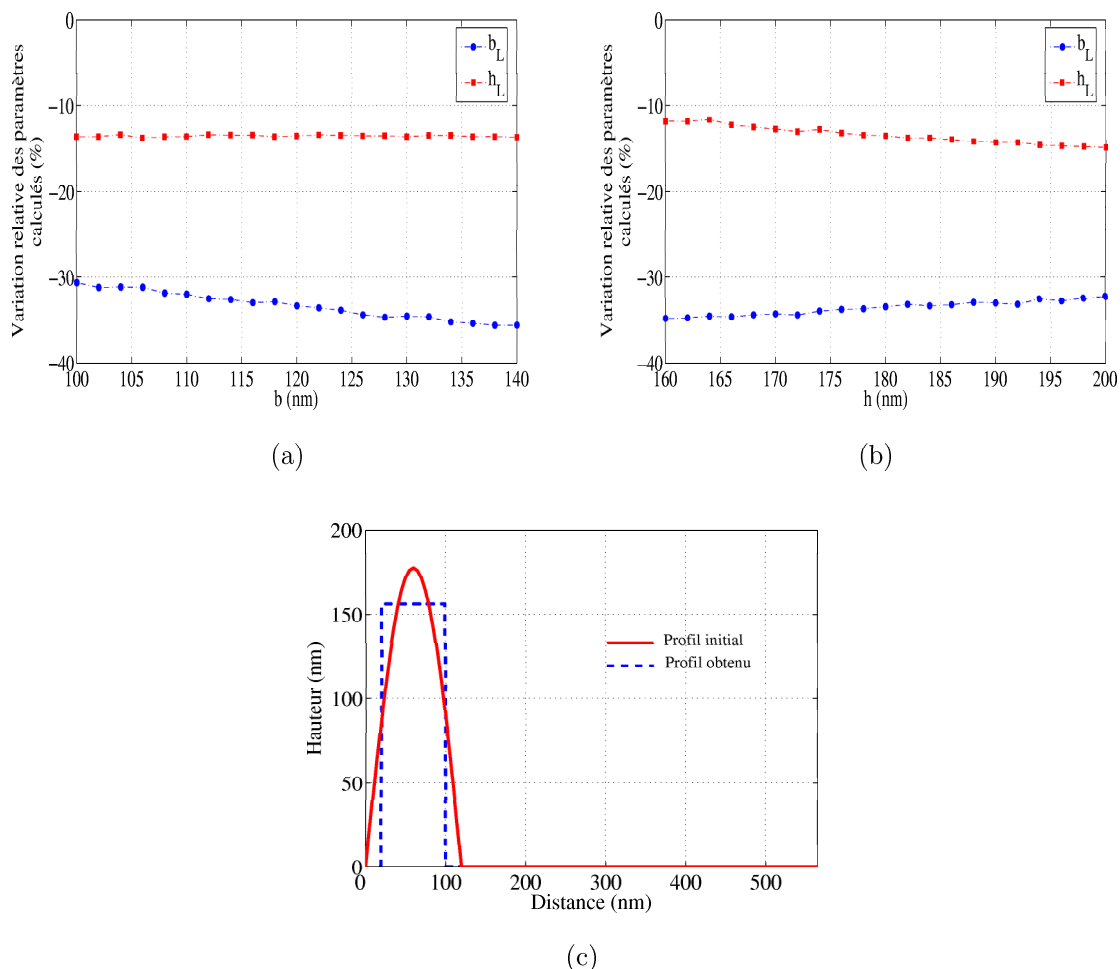


FIG. III.38: Variation relative des paramètres géométriques b_L et h_L du profil rectangulaire, par rapport à leurs valeurs nominales à savoir $b_L = b$ et $h_L = h$ en fonction de (a) b lorsque $h = 180$ nm et (b) h lorsque $b = 120$ nm. (c) Exemple d'un profil rectangulaire ($b_L = 80$ nm et $h_L = 155.5$ nm) obtenu par PMC à partir d'une signature ellipsométrique calculée pour un profil sinusoïdal ($b = 120$ nm et $h = 180$ nm).

par rapport aux valeurs nominales de b_s et h_s sont présentées sur les figures III.38(a) et III.38(b). Nous constatons que le profil rectangulaire estimé par le PMC est tel que sa ligne b_L est inférieure à la base de la sinusoïde b ainsi que sa hauteur h_L par rapport à la hauteur h . Ainsi, le profil rectangulaire obtenu est toujours superposé au profil sinusoïdal initial. Ceci peut être visualisé sur la figure III.38(c) pour un cas particulier où les paramètres du profil sinusoïdal sont $b = 120$ nm et $h = 180$ nm et ceux du profil rectangulaire estimés par le PMC sont $b_L = 80$ nm, $h_L = 155.5$ nm.

Nous avons vu dans les deux premières études l'impact du choix d'un profil simple rectangulaire lorsque l'échantillon à caractériser est sinusoïdal ou trapézoïdal. Le PMC réagit d'une façon similaire et tend à estimer le profil rectangulaire de la même façon dans les deux cas. Un écart non négligeable est obtenu entre le profil réel et le profil estimé par PMC.

P_1 : Trapèze arrondi en haut et en bas / P_2 : Trapèze - En pratique, il est impossible d'obtenir un profil géométrique parfait et ceci est vrai pour toutes les techniques

de fabrication. Si le profil trapézoïdal est le plus employé pour modéliser ceux obtenus expérimentalement, la présence des bords arrondis en haut ou en bas des sillons est généralement visible sur les images MEB. Dans ce paragraphe, nous allons examiner le cas où l'on doit caractériser un réseau de diffraction ayant un profil trapézoïdal à bords arrondis, en considérant que le profil est trapézoïdal simple.

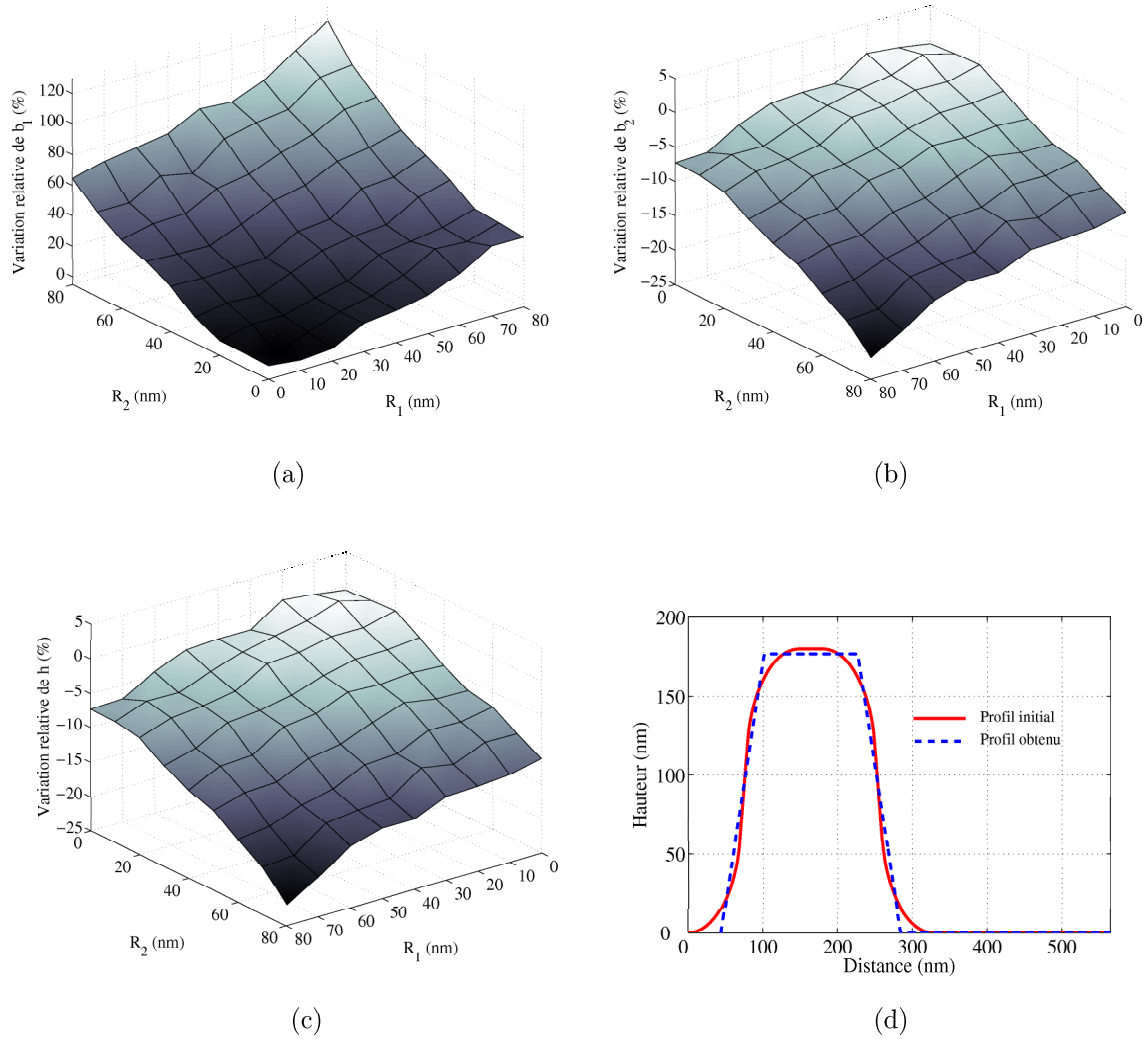


FIG. III.39: Variation relative des paramètres géométriques (a) b_1 , (b) b_2 et (c) h calculés par le PMC, par rapport à leurs valeurs nominales ($b_1 = 30$ nm, $b_2 = 120$ nm et $h = 180$ nm), en fonction des rayons de courbures bas R_1 et haut R_2 . (d) Exemple d'un profil trapézoïdal ($b_1 = 57.7$ nm, $b_2 = 124.9$ nm, $h = 176.9$ nm) obtenu par le PMC à partir d'une signature ellipsométrique correspondante à un profil trapézoïdal arrondi ($b_1 = 30$ nm, $b_2 = 150$ nm, $h = 180$ nm, $R_1 = 70$ nm, $R_2 = 70$ nm.)

Reprenons le PMC employé dans la partie III.2.1 entraîné pour estimer en sorties les paramètres géométriques correspondant à un profil trapézoïdal symétrique (b_1 , b_2 et h). Nous avons simulé les signatures ellipsométriques de réseaux de diffraction de profil trapézoïdal arrondi ayant $b_1 = 30$ nm, $b_2 = 120$ nm et $h = 180$ nm, en fonction des rayons de courbure R_1 et R_2 qui varient de 0 nm à 80 nm. Ces dernières sont introduites en entrée du PMC, les variations relatives des paramètres géométriques calculées en sortie, pour le profil trapézoïdal, sont tracées sur les figures III.39(a)-III.39(c).

Ces courbes montrent que la pente b_1 du profil estimé a tendance à augmenter lorsque les rayons de courbure des arrondis haut et bas augmentent, contrairement à la largeur de ligne b_2 qui diminue dans un ordre de grandeur moindre. Quant à l'épaisseur h , elle ne semble pas ou peu affectée. On décèle cependant une diminution notable lorsque l'arrondi en bas est de plus en plus grand. Ainsi, le profil estimé se trouve inclu dans le profil réel du réseau de diffraction, comme on le montre pour un cas particulier dans la figure III.39(d). Cela représente le cas où le PMC estime en sortie les paramètres $b_1 = 57.7$ nm, $b_2 = 124.9$ nm et $h = 176.9$ nm correspondant à un profil trapézoïdal, lorsque la signature introduite en entrée correspond à un profil arrondi de paramètres $b_1 = 30$ nm, $b_2 = 150$ nm, $h = 180$ nm, $R_1 = 70$ nm et $R_2 = 70$ nm. Il semble donc que dans le cas où l'on néglige des paramètres de second ordre dans la modélisation géométrique du profil, le PMC réalise une approximation du profil de la structure à partir de ses compétences géométriques ce qui n'était pas le cas dans les exemples précédents. Dans le même ordre d'idée, le paramètre h représentant la même quantité physique dans les 2 modèles ne se trouve que peu perturbé.

P1 : Trapèze symétrique avec couche résiduelle / P2 : Rectangulaire avec couche résiduelle - Dans certaines techniques de lithographie comme la nanoimpression, l'homogénéité de l'épaisseur de la couche résiduelle est un point clef afin d'assurer une gravure uniforme sur le substrat. Pour des raisons évidentes de simplicité, nous avons généralement recours à une modélisation géométrique de type profil rectangulaire simple avec une couche résiduelle (P_2). Ce dernier permet notamment de réduire considérablement le temps du calcul direct utilisé dans l'étape de résolution de problème inverse.

Nous allons étudier les conséquences de ce choix lorsque le profil réel au dessus de la couche résiduelle est un trapèze (P_1), ce qui est souvent le cas pour cette technique. Le réseau de diffraction employé est cette fois-ci en résine et déposée sur un substrat de silicium, de période $\Lambda = 400$ nm.

Le PMC est entraîné pour estimer les paramètres géométriques b , h et h_r du profil P_2 , compris dans les domaines suivants :

$$\begin{aligned} 250 < b < 350 \text{ nm} \\ 170 < h < 350 \text{ nm} \\ 0 < h_r < 80 \text{ nm} \end{aligned} \tag{III.13}$$

Ensuite, une série de signatures ellipsométriques calculées selon le modèle P_1 pour les échantillons suivants : $2 \text{ nm} < b_1 < 60 \text{ nm}$ (par pas de 2 nm), $b_2 = 280$ nm, $h = 250$ nm et $h_r = 40$ nm. La figure III.40 montre la variation relative des paramètres calculés par le PMC par rapport aux valeurs nominales fixées, lorsque la projection de la pente b_1 croît.

L'épaisseur h du réseau demeure presque inchangée lorsque b_1 est inférieur à 20 nm, tandis que les deux autres paramètres b_2 et h_r tendent à croître. Lorsque le profil réel possède une grande pente b_1 , les paramètres calculés par le PMC s'écartent considérablement des paramètres exactes. Les allures des courbes concernant les variations de b_2 et h rappelle celles des courbes III.37(a) en moindre mesure. En effet, ici c'est la couche résiduelle qui compense le plus les défauts de modélisation du profil par une forte augmentation. Ainsi, le choix d'un profil simple afin d'accélérer le temps de calcul est fortement déconseillé surtout lorsque la précision sur la couche résiduelle est une exigence.

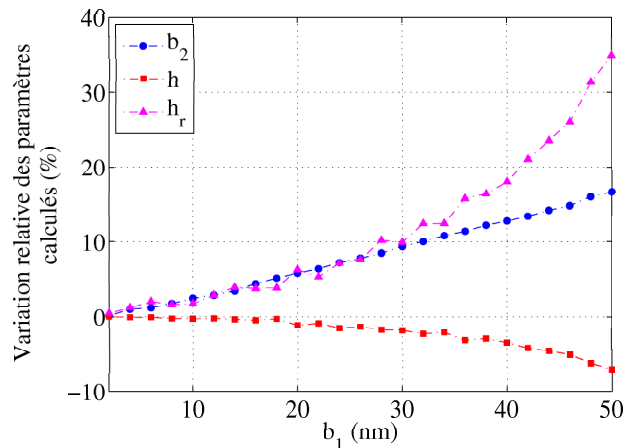


FIG. III.40: (a) Variation relative des paramètres géométriques b , h et h_r calculés par le PMC, par rapport à leurs valeurs initiales ($b = 280$ nm, $h = 250$ nm et $h_r = 40$ nm), en fonction de b_1 .

L'étude paramétrique que nous venons de présenter nous a permis de comprendre l'influence de certains paramètres internes ainsi que l'influence des hypothèses structurales sur les résultats du PMC. Cette étude a ainsi mis en évidence l'importance de connaître certaines conditions expérimentales de mesure avec précision, comme par exemple l'angle d'azimut φ . La période est également un paramètre à connaître avec précision. Nous avons vu qu'un choix incorrect de profil géométrique peut engendrer des erreurs remarquables sur les dimensions calculées par le PMC. Nous avons ainsi vu que dans la majorité des cas, le PMC donne une réponse comprise dans son domaine de validité même si l'hypothèse du profil géométrique est fausse. En effet, le PMC donne comme résultat le profil géométrique de sa base d'entraînement dont la signature est la plus proche de celle introduite en entrée et correspondant à un autre profil. Il faut noter que la seule façon de vérifier la véracité des résultats est de réaliser un calcul direct sur un ensemble test non utilisé lors de la phase d'entraînement du PMC et de déterminer l'erreur quadratique moyenne entre la signature calculée et celle mesurée.

III.4 Sélection de la signature ellipsométrique optimale

Il existe dans la littérature différents types de signatures optiques exploitées pour la caractérisation d'un réseau de diffraction. La mesure dans l'ordre 0 est sans aucun doute la plus utilisée lié certainement au fait que ce dernier existe toujours quelque soit la période et la longueur d'onde (paragraphe I.2.2.2 page 45). Le choix de la signature optique n'est pas anodin et influe grandement sur les résultats. La difficulté réside justement dans la sélection des mesures les plus pertinentes et les plus riches en information.

La signature optique peut être composée de mesures à différentes longueurs d'onde λ , à différents angles d'incidence θ ou même à différents angles d'azimut φ .

En scatterométrie, il est commun d'exploiter un seul type de signature optique lors de la caractérisation d'un réseau de diffraction. La plupart des auteurs utilisateurs de cette technique se sont rarement souciés du calcul de sensibilité puisque la signature employée répondait à leurs besoins. Toutefois, la diminution continue des dimensions fabriqués dans les dernières années a poussé quelques-uns à explorer davantage cette voie [174]. En effet,

il devient de plus en plus indispensable de sélectionner les entrées les plus sensibles dans le but de garder toujours une précision élevée et éliminer celles qui ne le sont pas afin d'éviter que ces dernières ne perturbent le calcul et affectent la précision souhaitée sur le résultat.

Dans cette partie, nous proposerons d'appliquer une méthode neuronale (la méthode HVS) afin de sélectionner les entrées de la signature optique les plus performantes.

III.4.1 Estimation des sensibilités de la signature optique

L'un des avantages pour la scatterométrie est la sensibilité de la signature optique diffractée par une structure périodique par rapport aux paramètres constitutifs de cette dernière, notamment ceux définissant sa géométrie. Cette sensibilité est définie comme étant une mesure de variation de la signature optique compte tenu d'un changement donné des paramètres géométriques. L'estimation de la sensibilité permet de déterminer les variables de mesures les plus influentes pour la caractérisation d'une structure donnée.

L'avantage de la technique ellipsométrique est le fait de pouvoir récolter un grand nombre de mesures selon les différentes configurations expérimentales, et de permettre d'accéder aussi à la phase. De même, l'acquisition des mesures devient de plus en plus rapide avec les avancées réalisées au niveau de l'automatisation des appareils. Il faut noter que la dimension de l'espace des observations peut être très vaste dans certains types de mesures. Ainsi, en ellipsométrie les intensités I_s et I_c peuvent être mesurées en fonction de plusieurs paramètres (angle d'incidence θ , longueur d'onde λ , angle d'azimut φ , etc. .).

Ce type d'étude est encore rare dans la littérature. Parmi les méthodes les plus répandues, on peut citer celles basées sur le calcul simple de la dérivée de la signature optique par rapport aux paramètres géométriques de la structure [175, 176]. Des études [177] ont montré l'importance d'éviter la corrélation entre les paramètres structurels. Dans ce contexte, certains auteurs ont proposé de réduire l'erreur induite par la corrélation entre les paramètres géométriques de la structure en utilisant un RN [178]. Une autre étude traitant du sujet et basée sur la sensibilité et les corrélations des paramètres géométriques est apparue récemment [164].

Nous allons illustrer, selon le principe simple du calcul de dérivées, le mode de sélection des points de mesure les plus pertinentes. Pour cela, considérons le réseau de diffraction en silice de période $\Lambda = 565$ nm et de profil supposé trapézoïdal. Les figures III.41(a) et III.41(c) montrent l'influence de la variation du paramètre géométrique b_2 sur la signature optique composée des intensités I_s et I_c pour $\theta = 60^\circ$ et $\varphi = 0^\circ$ en fonction de la longueur d'onde λ , pour deux réseaux dont les deux autres paramètres b_1 et h diffèrent.

Le calcul des dérivées par rapport à b_2 pour une longueur d'onde λ fixe pour les deux structures est représenté par les figures III.41(b) et III.41(d).

Ce calcul permet de localiser les régions spectroscopiques les plus sensibles, désignées par les fortes valeurs des dérivées. Ainsi, nous pouvons constater que ces régions ne sont pas identiques pour les deux réseaux étudiés. Pour le premier, il semble plus intéressant de travailler dans la région des faibles longueurs d'onde (300 nm-500 nm), tandis que le deuxième nécessite le choix des longueurs d'onde plus élevées (entre 600 nm et 700 nm).

De la même façon, nous allons étudier la sensibilité de la signature ellipsométrique par rapport à la hauteur h . Pour cela, considérons à nouveau deux réseaux de diffraction différents dont les variations des signatures ellipsométriques par rapport au paramètre h sont retracées sur les figures III.42(a) et III.42(c).

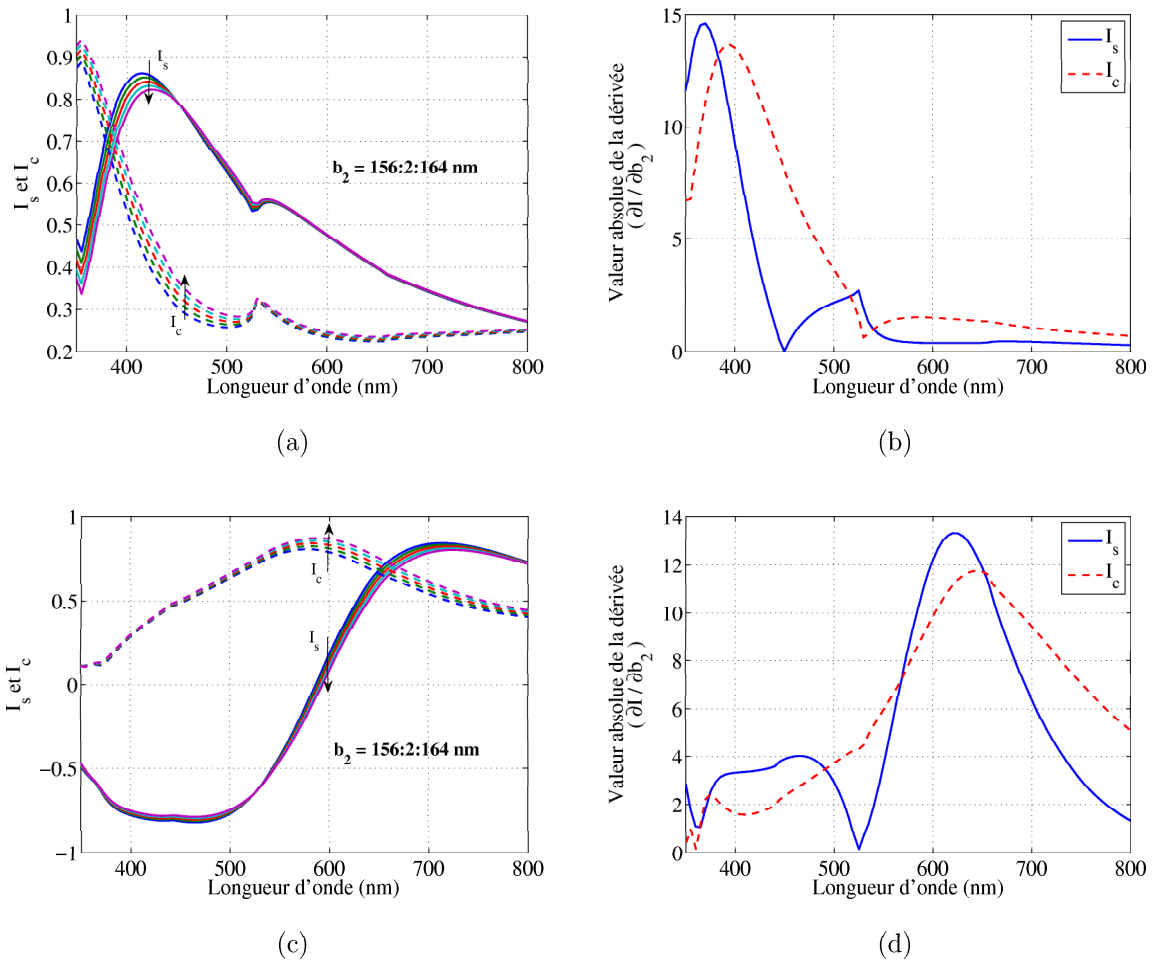


FIG. III.41: Variation des intensités I_s et I_c en fonction de la longueur d'onde λ pour plusieurs valeurs données de b_2 , lorsque (a) $b_1 = 30$ nm et $h = 120$ nm (c) $b_1 = 60$ nm et $h = 220$ nm. Les angles d'incidence et d'azimut sont fixés $\theta = 60^\circ$ et $\varphi = 0^\circ$. (b) et (d) représentent les valeurs absolues des dérivées de I_s et I_c par rapport à b_2 en fonction de la longueur d'onde, calculées pour les valeurs prises respectivement en (a) et en (c).

A partir des dérivées représentées sur les figures III.42(b) et III.42(d), nous remarquons à nouveau que les régions de fortes sensibilités trouvées ne correspondent pas exactement aux mêmes domaines de longueur d'onde.

Cette constatation est critique pour définir une méthode générique car les sensibilités semblent être différentes d'une structure à l'autre. Ce type de calcul peut être considéré comme local ; c'est à dire que la sensibilité est estimée pour un seul échantillon ayant des paramètres donnés. Il est difficilement généralisable si ce n'est dans une région limitée autour des valeurs nominales des paramètres géométriques étudiés. De plus, nous n'avons étudié ici qu'un cas simple correspondant à la variation d'un seul paramètre lorsque les autres sont gardés fixes, pour une signature spectroscopique. Or, il n'est pas évident que pour un autre jeu de paramètres fixes, les constatations soient les mêmes.

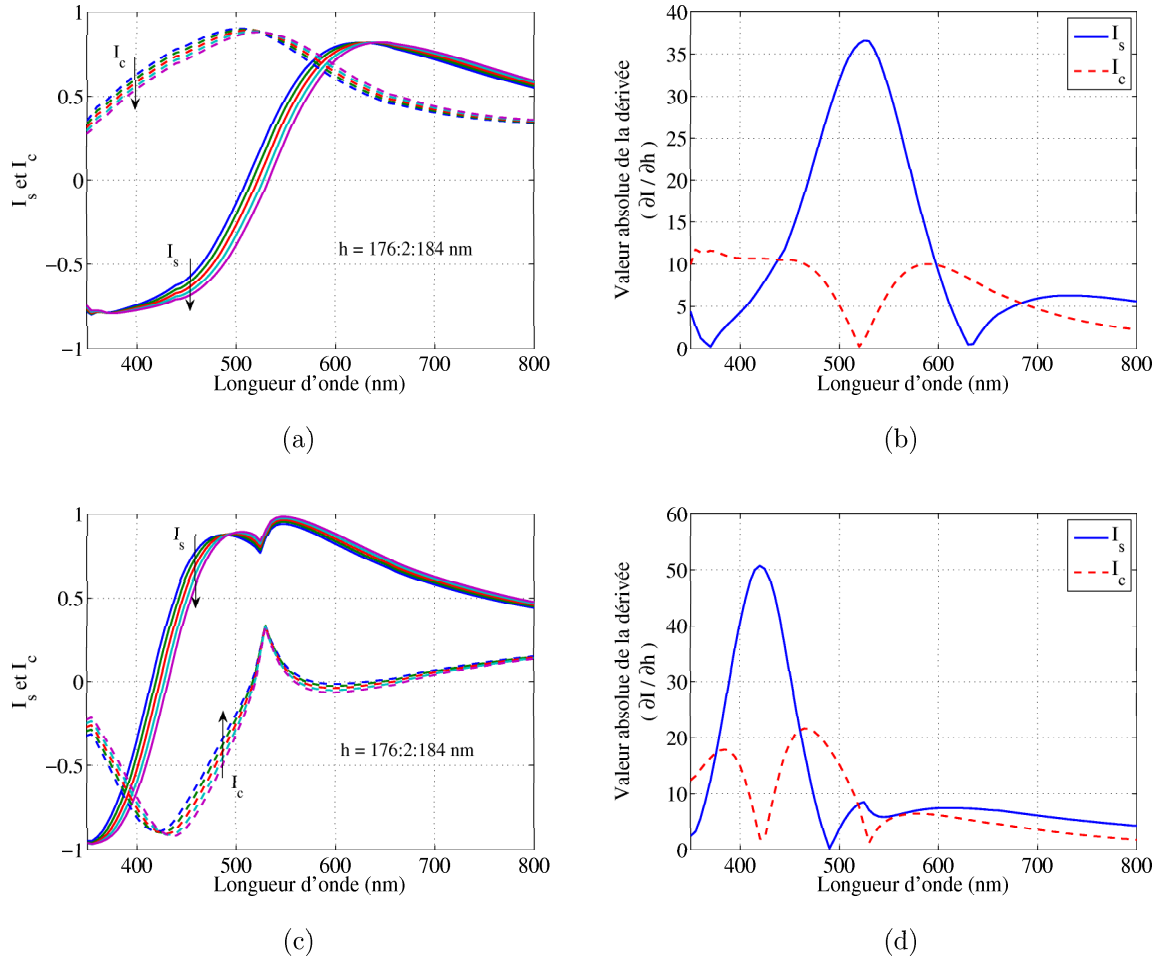


FIG. III.42: Intensités I_s et I_c en fonction de la longueur d'onde pour plusieurs valeurs données de h , lorsque (a) $b_1 = 20$ nm et $b_2 = 200$ nm (c) $b_1 = 40$ nm et $b_2 = 90$ nm. Les angles d'incidence et d'azimut sont fixés $\theta = 60^\circ$ et $\varphi = 0^\circ$. (b) et (d) représentent les valeurs absolues des dérivées de I_s et I_c par rapport à h en fonction de λ , calculées pour les valeurs prises respectivement en (a) et en (c).

III.4.2 Apport de la méthode neuronale

Dans le domaine de la scatterométrie utilisant des réseaux de neurones, la méthode HVS (voir II.6.2 page 112) a été appliquée pour la première fois par Robert et al. [174]. Cette méthode a permis de réduire le nombre des intensités diffractées utilisées pour la caractérisation, ainsi que la complexité du PMC employé.

Dans la présente étude, nous allons appliquer la méthode HVS pour la caractérisation du même réseau de diffraction, noté E_{hr}^1 , employé dans la partie III.2.4.3 page 153. Ce dernier est un réseau en résine de profil trapézoïdal symétrique possédant une couche résiduelle et déposé sur un substrat de silicium. Rappelons que la signature ellipsométrique est composée de 46 mesures, et le PMC de départ contient 20 neurones dans la couche cachée. La valeur de la période est fixée à $\Lambda = 675$ nm, telle qu'elle a été déterminée auparavant.

La figure III.43 représente l'évolution de l'erreur quadratique E_{test} en fonction du nombre d'entrée N lorsque la méthode HVS est employée pour la sélection des entrées.

Notons que beaucoup de mesures peuvent être supprimées tout en gardant une erreur

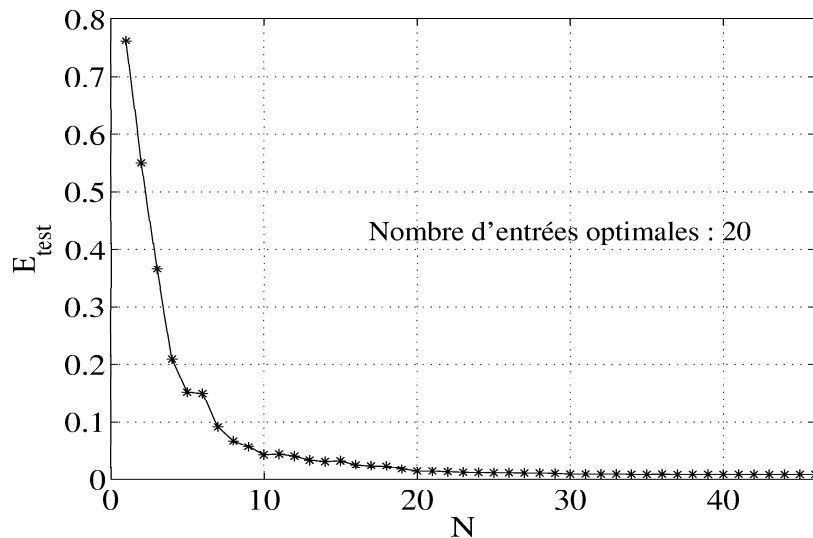


FIG. III.43: Evolution de l'erreur quadratique moyenne E_{test} en fonction du nombre d'entrées N .

E_{test} constante. D'après la courbe, nous avons retenu 20 entrées (voir paragraphe II.6.3 page 113) considérées comme les plus influentes et qui constitueront la nouvelle signature ellipsométrique réduite de moitié. Les entrées sélectionnées sont regroupées dans le tableau III.12, ainsi que leur niveau de pertinence.

La sélection du nombre de neurones N_c est ensuite réalisée à partir de la courbe de la figure III.44. Le nombre retenu est $N_c = 12$, permettant ainsi de garder les mêmes performances du PMC initial supposé représenter le modèle global. Un nouvel entraînement est effectué avec la nouvelle architecture retenue. Le réseau E_{hr}^1 est de nouveau caractérisé. Les paramètres calculés après sélection HVS sont comparés aux résultats donnés par le PMC initial (tableau III.5 page 154), dans le tableau III.13.

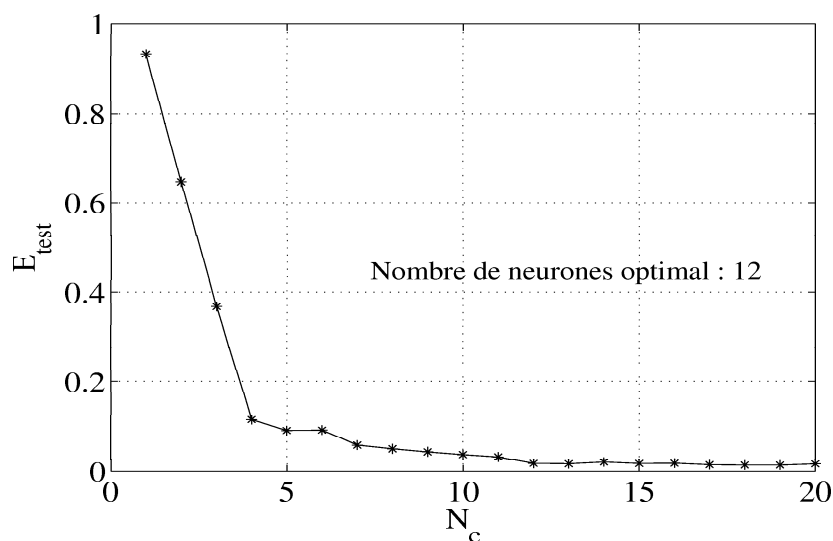


FIG. III.44: Evolution de l'erreur quadratique moyenne E_{test} en fonction du nombre de neurones cachés N_c .

Les valeurs des paramètres géométriques montrent une concordance entre les résultats

$\lambda(nm)$	350	370	390	410	430	450	470	490	510	530	550	570
I_s			● ₂				● ₃	● ₁₇	● ₆	● ₁₈	● ₁₀	● ₈
I_c			● ₅					● ₉				
$\lambda(nm)$	590	610	630	650	670	690	710	730	750	770	790	
I_s	● ₁₄	● ₁₃		● ₁	● ₁₆	● ₁₁	● ₁₂	● ₁₅	● ₂₀	● ₇	● ₁₉	
I_c											● ₄	

TAB. III.12: Liste des intensités (I_s , I_c) sélectionnées par la méthode HVS. L'importance de chacune d'elle est donnée en indice.

obtenus par le PMC simplifié après HVS et le PMC entier de départ. Les mêmes performances sont obtenues avec un PMC moins complexe et une signature ellipsométrique réduite.

	Paramètres géométriques calculés par PMC			
	$b_1(nm)$	$b_2(nm)$	$h(nm)$	$h_r(nm)$
sans HVS	69,3±8,3	51,4±1,2	252,6±4,1	520,5±1,5
avec HVS	67,1±5,3	52,9±3,5	250,4±3,8	521,2±2

TAB. III.13: Comparaison entre les paramètres géométriques du profil de réseau, noté $E_{h_r}^1$, estimés par le PMC avant et après sélection HVS.

Finalement, nous allons utiliser le PMC simplifié pour contrôler l'homogénéité de l'échantillon. Les différents points de mesure sont montrés sur le schéma de la figure III.45(a). Nous avons considéré 3 lignes (L_1 , L_2 et L_3) espacées de 0.5 cm, et sur chacune d'elles nous avons effectué 8 mesures équidistantes. Une caractérisation MEB est également effectuée au centre du réseau sur une surface de dimension $30 \mu m \times 40 \mu m$, donnée par la figure III.45(b). Cette dernière montre que le réseau ne présente pas de défauts remarquables localement.

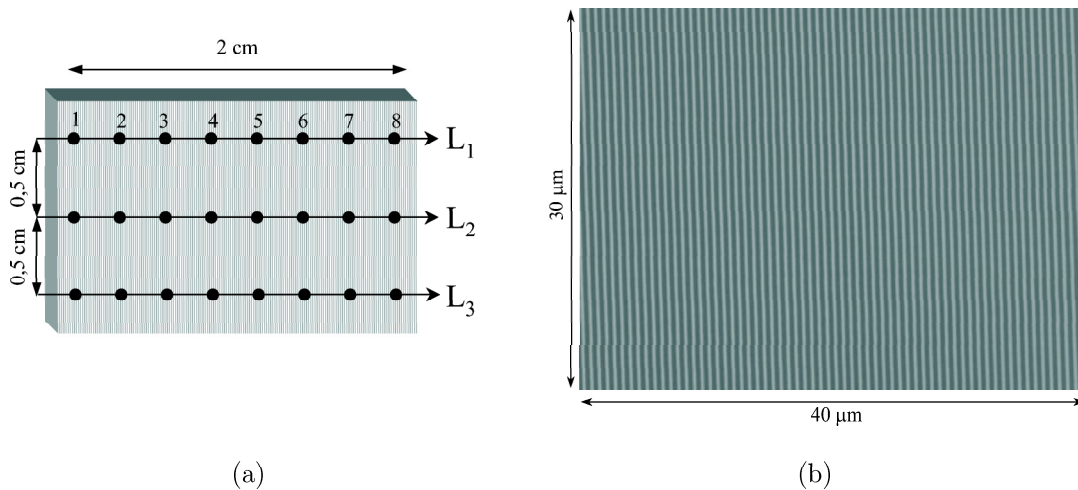


FIG. III.45: (a) Points de mesure choisis pour effectuer une cartographie de l'échantillon $E_{h_r}^1$ sur une dimension de $2 \text{ cm} \times 1 \text{ cm}$, ainsi que (b) l'image MEB du même échantillon sur une surface de dimension $30 \mu m \times 40 \mu m$.

Le résultat de la cartographie à l'aide du PMC simplifié est comparé à celui obtenu par le PMC complet, comme le montre la figure III.46.

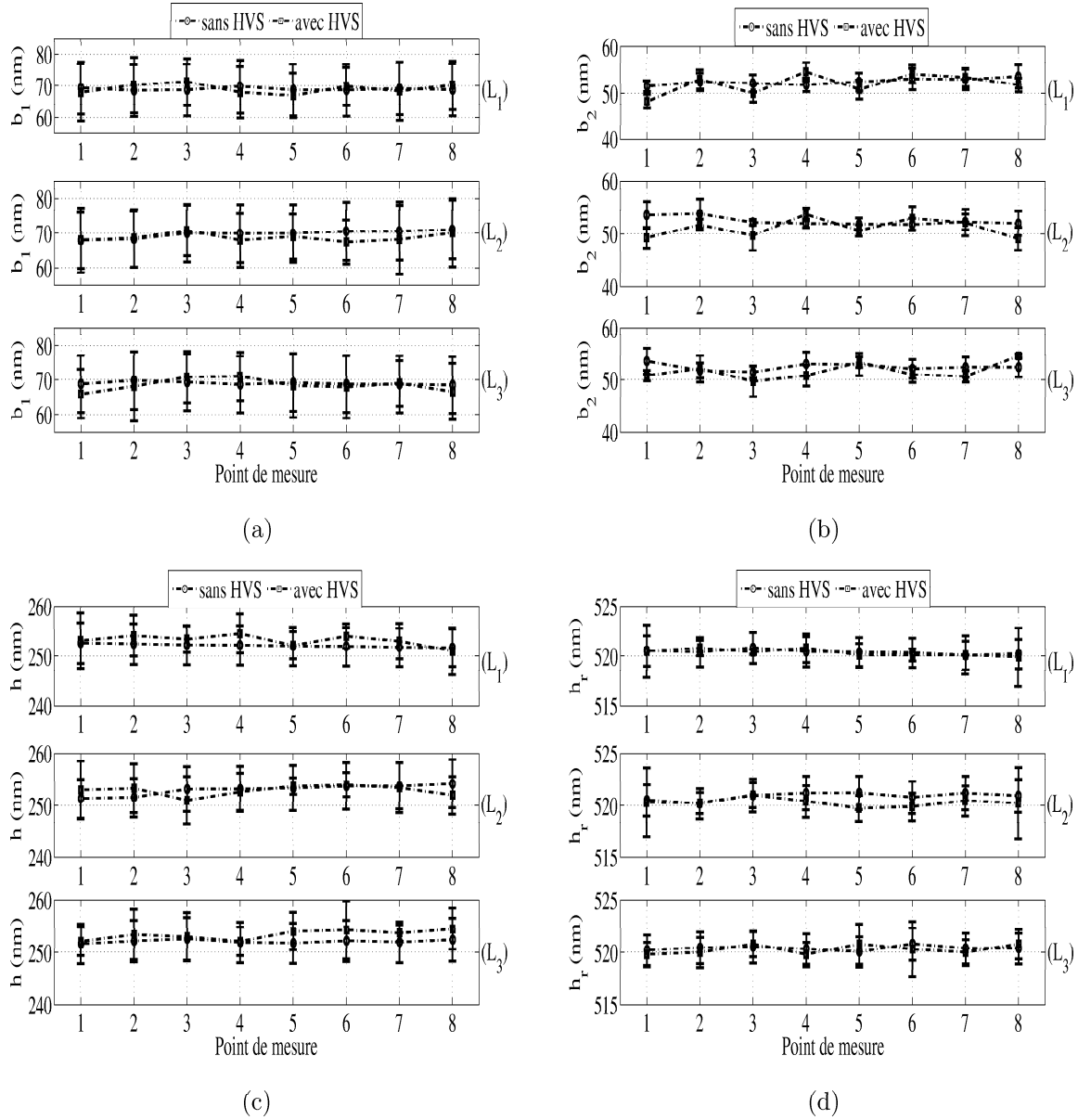


FIG. III.46: Cartographie des paramètres géométriques (a) b_1 , (b) b_2 , (c) h et (d) h_r sur une large surface de l'échantillon ($2 \text{ cm} \times 1 \text{ cm}$) E_{hr}^1 obtenue par caractérisation neuronale avec signature complète et avec signature optimale sélectionnée par HVS.

Ces courbes nous révèlent que l'échantillon est parfaitement homogène sur une large surface. Les résultats donnés par le PMC simplifié après HVS concordent bien avec ceux donnés par le PMC complet. Nous pouvons conclure que le procédé de fabrication a été bien maîtrisé permettant d'obtenir de réseaux de haute qualité géométrique.

Nous venons de présenter la méthode HVS appliquée dans le domaine de la scatterométrie. Elle permet de sélectionner la signature optique optimale dans un espace donné des paramètres géométriques, qui peut être plus ou moins grand. La complexité du RN est ainsi réduite par un choix de nombre de neurones cachés optimal. La durée globale de la caractérisation (mesure et traitement) se trouve ainsi considérablement réduite. Ceci nous a permis de contrôler d'une façon rapide et précise l'homogénéité de nos échantillons sur de larges surfaces.

III.5 Identification des profils géométriques

Rappelons qu'en scatterométrie la résolution du problème inverse nécessite la connaissance *a priori* du profil géométrique de la structure à caractériser. Nous avons montré dans la partie III.3.2.3 page 166 que, dans le cas où le profil supposé est plus ou moins différent du profil réel, ce problème est mal posé et les conséquences peuvent être majeures sur les résultats.

A notre connaissance, il n'existe pas aujourd'hui dans la littérature scientifique une méthode qui permette d'identifier sans connaissance *a priori* le profil géométrique supposé pour une structure donnée. Généralement, la vérification de l'hypothèse sous jacente du profil a lieu après la phase de caractérisation, par estimation de la qualité du résultat obtenu au moyen du calcul direct (voir par exemple partie III.2.2.2 figure III.16). Si le résultat obtenu n'est pas satisfaisant, une modification du profil géométrique est nécessaire et indispensable. Les seules études récentes [179] qui permettent d'effectuer une caractérisation sans connaissance *a priori* du profil, sont basées sur la reconstruction de la surface du réseau à partir d'une mesure de phase et d'amplitude du champ diffracté. Toutefois, ces méthodes utilisant l'approche scalaire ne peuvent pas être appliquées pour la caractérisation de réseaux de faible période.

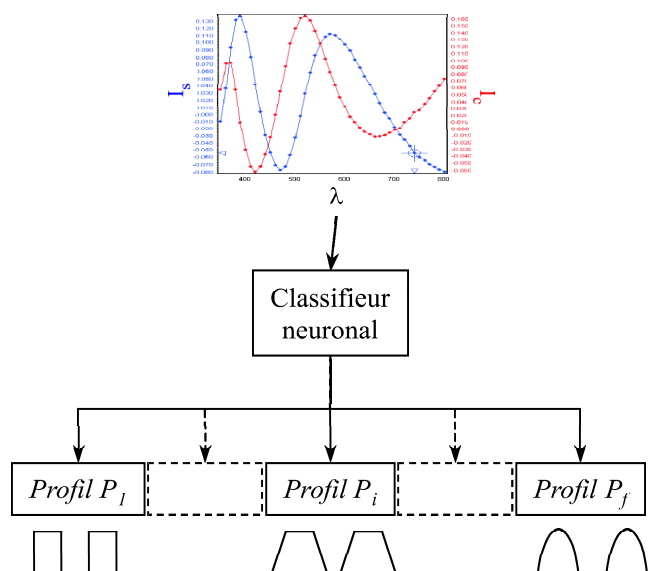


FIG. III.47: Schéma illustrant le principe du classifieur employé pour l'identification du profil géométrique d'un réseau de diffraction à partir de sa signature ellipsométrique mesurée.

Dans ce cadre, nous allons proposer une nouvelle approche pour identifier le profil géométrique d'un réseau de diffraction à partir de sa seule signature ellipsométrique mesurée. Cette approche emploie la classification neuronale décrite précédemment dans le manuscrit (voir partie II.7 page 116). Ainsi, un classifieur neuronal est conçu pour estimer en sortie les probabilités *a posteriori* d'appartenance aux classes représentant les divers profils géométriques considérés, notés $(P_1, \dots, P_i, \dots, P_f)$ (figure III.47). Ceci concerne un réseau de diffraction donné dont la signature mesurée est introduite en entrée du classifieur. Rappelons que la classification est supervisée, c'est-à-dire que le type et le nombre de profils parmi lesquels le RN doit choisir sont connus à l'avance et fournis durant l'apprentissage.

En général, les performances du classifieur sont évaluées par le pourcentage de classe-

ments corrects et incorrects. Rappelons que le classifieur ne prend pas de décision mais estime les probabilités a posteriori. Ensuite, la règle de décision de Bayes est généralement employée.

Nous allons à présent adopter un critère de décision spécifique. Ce dernier consiste à fixer une probabilité, appelée probabilité seuil, notée p_s et égale par exemple à 0,68. Reprenons le cas général de la figure III.47. Dans ce cas, le vecteur de sortie calculé par le PMC à partir d'un vecteur d'entrée particulier appartenant à la classe i peut s'écrire $[s_1 \dots s_i \dots s_f]$. Deux configurations sont ainsi possibles :

- $s_i > p_s$: la classe d'appartenance choisie est i et la classification est correcte,
- $s_i < p_s$: l'entrée n'est pas affecté à la classe i et n'est pas bien classée dans ce cas.

Pour tester la qualité du classifieur entraîné, nous considérerons ainsi un ensemble d'échantillons non utilisés pendant la phase d'entraînement. Les performances seront données par trois quantités : le pourcentage de classification correcte notée PCC , le pourcentage de classification incorrecte noté PCI et le pourcentage de non classification PNC , obtenues après décision prise suivant la règle donnée ci dessus.

Les deux applications qui seront fournies pas la suite visent deux problématiques distinctes bien que le principe de résolution soit le même. La première est destinée à identifier le profil du réseau qui sera utilisé ensuite comme hypothèse sous jacente pour la résolution du problème inverse. La seconde est dédiée à la détection d'une couche résiduelle au fond d'un réseau en résine indésirable pour les techniques de lithographie standards.

III.5.1 Etude du réseau en silice

III.5.1.1 Validation sur des réseaux simulés

Le réseau de diffraction employé pour cette étude est en silice de période fixe $\Lambda = 565$ nm. Nous allons considérer le cas particulier de deux profils géométriques : trapézoïdal symétrique (figure I.14 page 52) et sinusoïdal (figure III.18 page 149).

Pour la mise en place du classifieur, nous devons d'abord générer les données d'entraînement. Dans ce cas, elles sont formées par un nombre égal de signatures ellipsométriques calculées pour des réseaux de diffraction simulés avec l'un et l'autre des profils choisis. Cette signature est composée de 46 intensités I_s et I_c calculées pour $\theta = 60^\circ$, $\varphi = 0^\circ$ et λ est compris entre 350 nm et 800 nm par pas de 20 nm. Les paramètres des réseaux simulés sont compris dans les domaines de variation suivants, pour le profil trapézoïdal :

$$\begin{aligned} 0 < b_1 &< 60 \text{ nm} \\ 60 < b_2 &< 160 \text{ nm} \\ 150 < h &< 220 \text{ nm} \end{aligned} \tag{III.14}$$

et pour le profil sinusoïdal :

$$\begin{aligned} 60 < b &< 280 \text{ nm} \\ 150 < h &< 220 \text{ nm} \end{aligned} \tag{III.15}$$

Comme le codage des classes doit être binaire, le vecteur de sortie du classifieur est de dimension égale au nombre de classes et noté $[s_1 \ s_2]$. La classe correspondante au profil trapézoïdal est codée par le vecteur $[1 \ 0]$, et $[0 \ 1]$ pour le profil sinusoïdal.

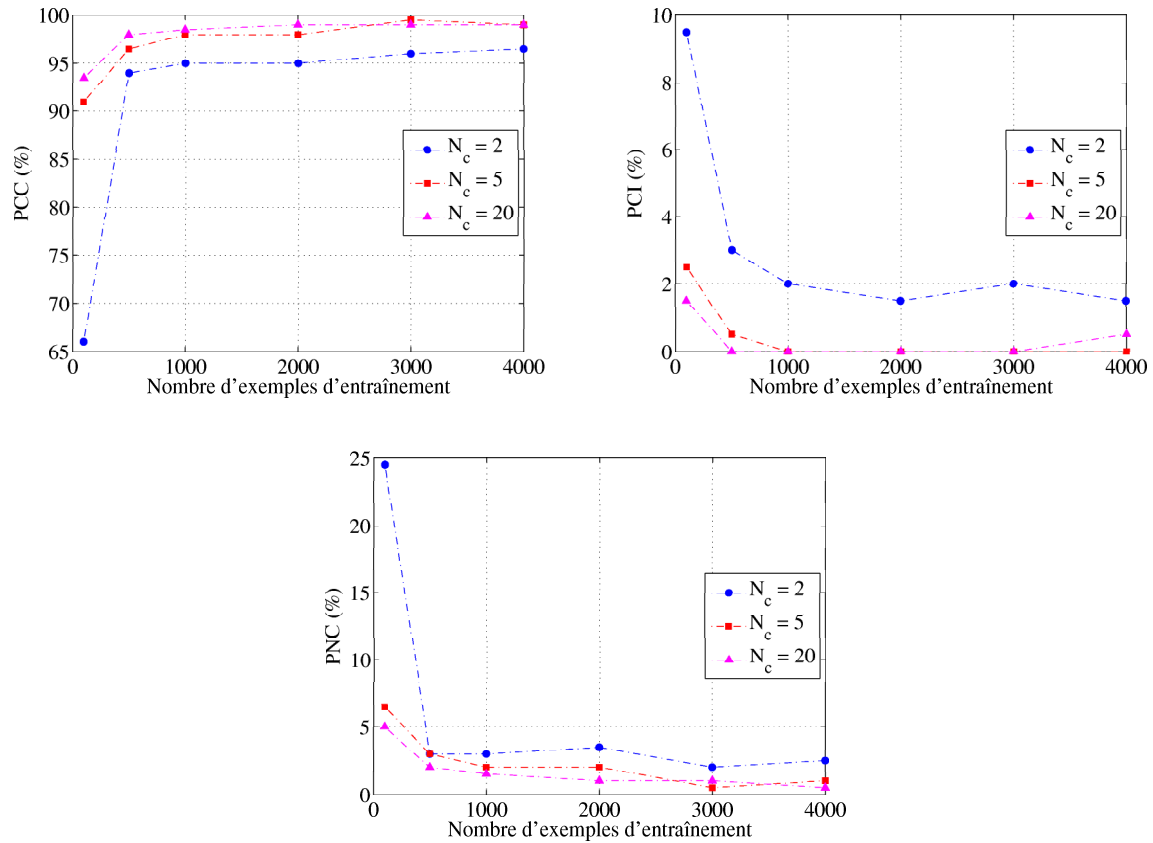


FIG. III.48: Evaluation des performances du classifieur en fonction du nombre d'exemples d'entraînement pour différents nombres de neurones : (a) pourcentage de classification correcte PCC (b) pourcentage de classification incorrecte PCI (c) pourcentage de non classification PNC .

Une fois l'entraînement réalisé, les pourcentages PCC , PCI et PNC sont estimés sur 200 réseaux simulés (100 de chaque corpus) pour tester les performances du classifieur. Le résultat est représenté sur les figures III.48, en fonction du nombre de couples d'entraînement n pour différents nombre N_c de neurones cachés. L'allure de ces courbes est identique. Le classifieur ne possède pas de bonnes performances avec un nombre faible de neurones, à savoir $N_c = 2$ dans ce cas. De plus, l'ajout de neurones cachés au delà d'un certain nombre ($N_c = 5$) est sans intérêt et n'améliore plus le résultat. De même, l'emploi d'un nombre faible d'exemples d'entraînement est insuffisant pour représenter correctement l'espace engendré par les deux classes. Les performances du classifieur tendent à se stabiliser à partir d'un nombre donné d'exemples. Dans ce cas 1000 couples s'avèrent suffisants.

Le fonctionnement du classifieur peut être validé sur deux échantillons particuliers simulés. Les réseaux choisis de profil sinusoïdal (P_1) et trapézoïdal (P_2) sont représentés sur la figure III.49(a), tel que les paramètres géométriques correspondants sont respectivement $b = 170$ nm, $h = 200$ nm pour P_1 et $b_1 = 10$ nm, $b_2 = 100$ nm, $h = 170$ nm pour P_2 .

La particularité de ces deux réseaux est qu'ils possèdent des signatures ellipsométriques voisines, tracées sur la figure III.49(b). Ceci montre que deux réseaux de géométries différentes peuvent avoir des signatures ellipsométriques presque similaires comme cela a été mis en évidence au paragraphe III.3.2.3 page 166. Il faut aussi noter que ces signatures

sont théoriques. Ainsi, dans des conditions expérimentales d'étude le bruit de mesure accentuera cette ambiguïté. Ce point peut constituer un vrai problème lorsqu'un de ces réseaux est caractérisé en supposant l'autre modèle, comme nous avons vu dans la partie III.3.2.3 page 166.

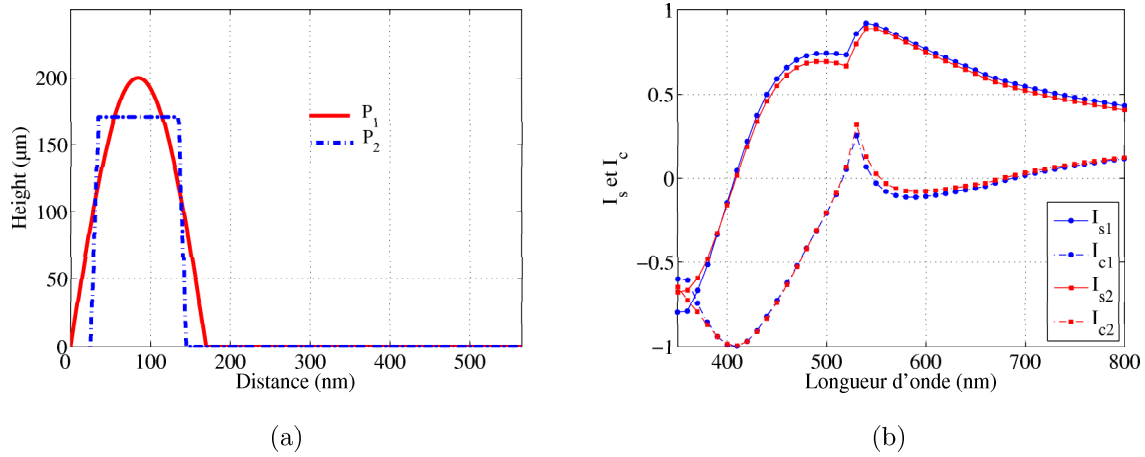


FIG. III.49: (a) Représentation des deux profils géométriques étudiés : P_1 sinusoïdal (traits continus) tel que $b = 170$ nm et $h = 200$ nm ; et P_2 trapézoïdal (traits discontinus) tel que $b_1 = 10$ nm, $b_2 = 100$ nm et $h = 170$ nm ; la période est de 565 nm. (b) Signatures ellipsométriques théoriques correspondantes (I_s, I_c en configuration classique $\varphi = 0^\circ$ tel que l'angle d'incidence $\theta = 60^\circ$) calculées dans le cas de deux réseaux de diffraction en silice.

Les signatures bruitées des deux réseaux de profils P_1 et P_2 sont introduites en entrée du classifieur. Le tableau III.14 regroupe les sorties, calculées par le PMC pour différents nombre d'exemples d'entraînement n . Dans le cas où $n = 100$, le résultat ne conduit pas à l'affectation d'une classe d'après le critère test choisi. La signature de chacun des deux réseaux est donc non classée. Ce nombre n est tout simplement insuffisant pour représenter l'espace des deux classes comme il est illustré par les courbes de la figure III.48. Les résultats corrects sont obtenus pour un nombre plus élevé d'exemples (dès $n = 2000$) assurant une bonne représentation dans l'espace considéré.

Réseau / Sorties du PMC	Nombre d'exemples d'entraînement n			
	100	2000	3000	4000
P_1 / s_1	0,41	0,19	0,13	0,13
P_1 / s_2	0,59	0,81	0,87	0,87
P_2 / s_1	0,54	0,99	0,99	0,99
P_2 / s_2	0,46	0,01	0,01	0,01

TAB. III.14: Valeurs des sorties s_1 et s_2 calculées par le classifieur entraîné avec $N_c = 5$ pour différents nombres d'exemples d'entraînement n , à partir des signatures ellipsométriques bruitées pour les deux réseaux P_1 et P_2 .

III.5.1.2 Résultat expérimental

La validation expérimentale du classifieur est réalisée sur le réseau en silice qui a été caractérisé dans la partie III.2.2. La signature ellipsométrique mesurée est introduite

en entrée du classifieur entraîné avec $N_c = 5$. Les sorties calculées par ce dernier sont regroupées dans le tableau III.15 pour différents nombre d'exemples d'entraînement n .

Sorties du PMC	Nombre d'exemples d'entraînement n			
	100	2000	3000	4000
s_1	0,03	0,79	0,89	0,92
s_2	0,97	0,21	0,11	0,08

TAB. III.15: Valeurs des sorties s_1 et s_2 calculées par le classifieur entraîné avec $N_c = 5$ pour différents nombres d'exemples d'entraînement n , à partir de la signature ellipsométrique mesurée sur le réseau en SiO_2 de période $\Lambda = 565$ nm.

Rappelons que l'échantillon considéré possède un profil plus proche d'un trapèze que d'un sinus, d'après l'image obtenue par MEB (figure III.13 page 144). Le résultat attendu est que le réseau appartienne à la classe représentant le profil trapézoïdal, codée par le vecteur $[1 \ 0]$. Le classifieur affecte au réseau un profil sinusoïdal avec une probabilité égale à 0,97 lorsque le nombre n est très faible. En effet, comme nous l'avons démontré précédemment, le nombre d'exemples n n'est pas suffisant pour représenter correctement l'espace d'étude et les résultats sont donc peu précis. Lorsque n augmente (à partir de $n = 2000$), les résultats sont corrects et cohérents avec les images MEB. Il faut noter que si la probabilité seuil p_s est fixée à une valeur proche de 0,9 il aurait fallu utiliser $n = 4000$ échantillons pour un résultat correct.

Nous venons de montrer une première application de la classification neuronale en scatterométrie, destinée à l'identification du profil géométrique d'un réseau de diffraction dans le cas particulier de deux profils sinusoïdal et trapézoïdal. Mais, il faut noter que la méthode peut être généralisée en incluant un nombre plus grand de profils. L'application suivante concerne la détection d'un défaut résultant d'une mauvaise fabrication de masque de résine.

III.5.2 Détection d'une couche résiduelle au fond du réseau

III.5.2.1 Validation théorique

Dans le même contexte que l'étude précédente, nous allons à présent concevoir un classifieur permettant de détecter une couche résiduelle au fond d'un réseau de diffraction. Cette étude entre dans le cadre de la détection de défaut lié à la présence de cette couche. Le but de cette opération est de pouvoir ainsi identifier rapidement l'existence de cette couche en sortie d'une chaîne de fabrication afin de juger ainsi la qualité de l'échantillon et décider ainsi l'arrêt éventuel avant la phase de gravure ultérieure. Notons que ce n'est pas le cas pour les réseaux obtenus par nanoimpression où cette dernière est laissée volontairement (paragraphe III.1.2). Nous avons ainsi opté pour des échantillons issus de la nanoimpression pour valider la méthode. Ainsi, le réseau de diffraction étudié est de période $\Lambda = 400$ nm constitué de résine déposée sur un substrat de silicium.

Les deux profils géométriques considérés sont le profil trapézoïdal simple (figure I.14) et le profil trapézoïdal avec une couche résiduelle au fond (figure III.22 page 153).

La signature ellipsométrique utilisée est composée de 46 mesures I_s et I_c , identiquement à celle prise dans l'étude précédente. Pour l'entraînement du classifieur, les données sont calculées à partir des domaines suivants définis pour le profil trapézoïdal par :

$$\begin{aligned}
0 < b_1 < 50 \text{ nm} \\
250 < b_2 < 350 \text{ nm} \\
170 < h < 350 \text{ nm}
\end{aligned}
\tag{III.16}$$

et pour le profil trapézoïdal avec couche résiduelle par :

$$\begin{aligned}
0 < b_1 < 50 \text{ nm} \\
250 < b_2 < 350 \text{ nm} \\
170 < h < 270 \text{ nm} \\
0 < h_r < 80 \text{ nm}
\end{aligned}
\tag{III.17}$$

Le codage des classes est tel que le vecteur $[1 \ 0]$ représente le profil trapézoïdal et le vecteur $[0 \ 1]$ le profil trapézoïdal avec une couche résiduelle.

L'entraînement réalisé, nous allons procéder maintenant à l'évaluation théorique du classifieur. Pour cela, nous avons généré un ensemble de 400 échantillons simulés équirépartis dans les deux classes (avec et sans couche). Les signatures correspondantes sont introduites en entrée du PMC classifieur. Nous avons représenté sur la figure III.50 la seule sortie s_1 (s_2 étant égale à $1 - s_1$). Pour faciliter la lecture de la courbe, les 200 premiers échantillons (axe des abscisses) correspondent au profil sans couche dont la valeur de s_1 attendue est de 1, les autres sont de profil avec une couche résiduelle et la valeur attendue est bien entendu 0.

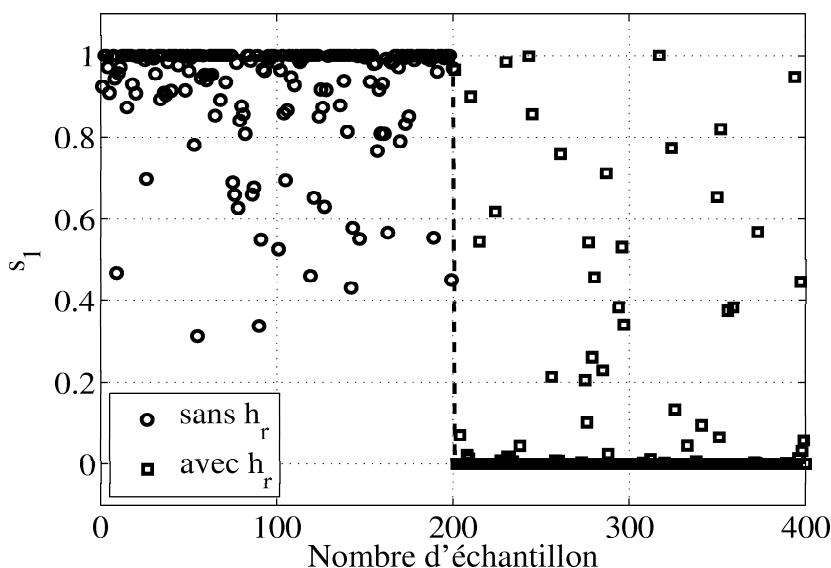


FIG. III.50: Sortie s_1 calculée par le classifieur pour un ensemble de 400 échantillons simulés équirépartis pour chaque profil (trapèze avec et sans couche résiduelle h_r). Les 200 premiers échantillons possèdent un profil trapézoïdal simple sans couche et les 200 autres un profil avec une couche résiduelle.

Dans le cas où p_s est prise égale à 0.68, le classifieur fonctionne plus ou moins bien en notant des erreurs de classification. Mais, les erreurs qui sont moins souhaitables sont celles de la partie de droite, c'est-à-dire les échantillons possédant une couche résiduelle et qui se sont classés dans l'autre classe. Intéressons nous plus particulièrement à cette partie de la figure III.50, à savoir les échantillons qui sont supposés appartenir à la classe

représentant le profil avec couche résiduelle. Nous allons essayer de quantifier la limite de notre classifieur, c'est-à-dire la plus petite épaisseur résiduelle détectable. Supposons que p_s soit égale 0,68. Cela revient à accepter un taux d'erreur de 32%. Ainsi, tous les échantillons dont la sortie s_1 est inférieur à 0,32 seront considérés comme bien classés. Les valeurs de h_r correspondantes aux échantillons ne vérifiant pas ce critère parmi les 200 échantillons test sont montrées sur la figure III.51(a).

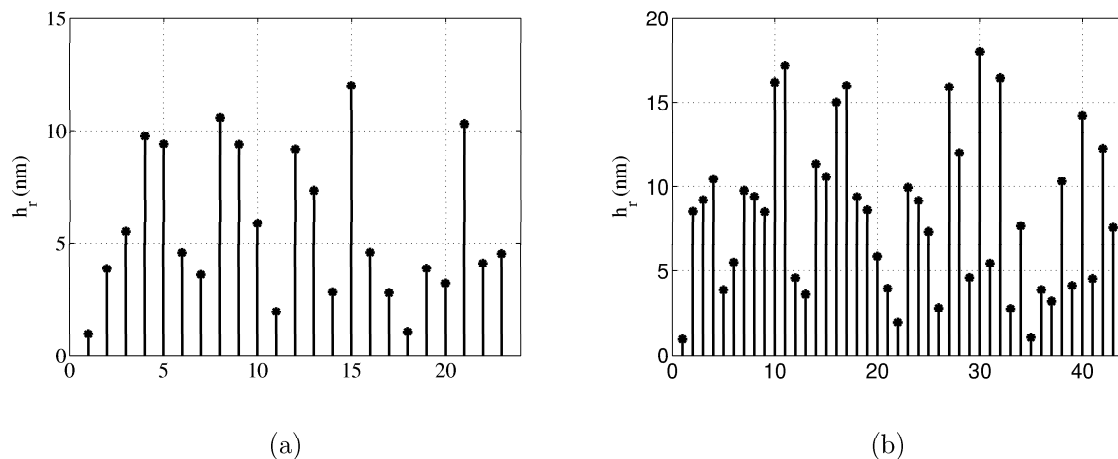


FIG. III.51: Valeurs des épaisseurs des couches résiduelles h_r correspondantes aux échantillons test de la classe avec couche résiduelle dont la sortie (a) $s_1 < 0,32$ et (b) $s_1 < 0,05$.

D'après ces valeurs, nous constatons que les échantillons dont l'épaisseur h_r dépasse environ 12 nm sont bien classés. Cela donne une idée claire de la capacité du classifieur ainsi qu'une estimation de sa limite théorique. Nous pouvons également réduire le taux d'erreur en augmentant la valeur de p_s . Si cette dernière est fixée à $p_s = 0,95$ par exemple (le taux d'erreur étant de 5%), les valeurs des épaisseurs des échantillons vérifiant ce critère plus sévère de classification sont données sur la figure III.51(b). La limite de détection est de l'ordre de 18 nm dans ce cas.

Notons cependant que cette limite dépend bien entendu de la signature ellipsométrique utilisé. Il est évident qu'une recherche de signature optimale permettrait de réduire cette limite. Cette étude nécessaire n'est pas traitée dans ce manuscrit mais s'inscrit dans la poursuite de ces travaux.

III.5.2.2 Résultat expérimental

Pour valider la méthode sur un cas réel, nous avons considéré un réseau de diffraction, noté $E_{h_r}^5$ issu d'un procédé de nanoimpression. Dans ce cas, la présence de la couche résiduelle est certaine. Une image MEB de cet échantillon est montrée sur la figure III.52.

La signature ellipsométrique mesurée sur ce dernier est introduite en entrée du classifieur. Le vecteur de sortie $[s_1 \ s_2]$ calculé par le PMC classifieur pour différents couples d'entraînement n est donné dans le tableau III.16.

Les résultats donnés dans ce tableau sont cohérents avec l'image MEB de l'échantillon. Les sorties estimées, même avec un nombre d'exemples n faible, correspondent bien à la classe attendue.

Pour faciliter le traitement par un second PMC, nous allons procéder à la caractérisation de l'échantillon. Pour ce faire, considérons un nouveau PMC entraîné selon le profil

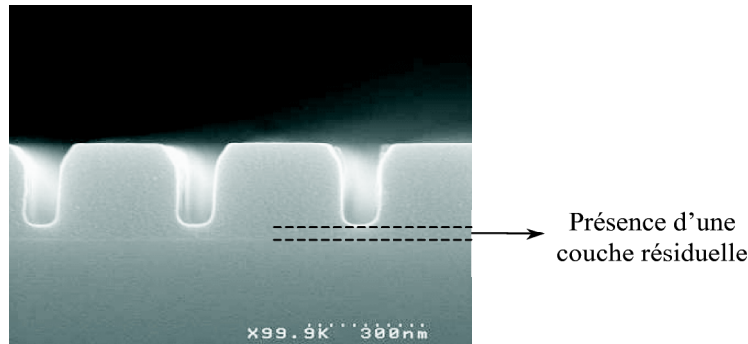


FIG. III.52: Image MEB de l'échantillon montrant la présence d'une couche résiduelle.

Sorties du PMC	Nombre de couples n		
	100	1500	5000
s_1	0,12	0,01	0
s_2	0,88	0,99	1

TAB. III.16: Valeurs des sorties s_1 et s_2 calculées par le classifieur entraîné avec $N_c = 5$ pour différents nombres d'exemples d'entraînement n , pour le réseau E_{hr}^5 .

trapézoïdal avec une couche résiduelle (figure III.22 page 153). Les paramètres géométriques varient selon le domaine défini par l'équation III.16. Le nombre de neurones étant égal à $N_c = 20$. Le résultat est donné dans le tableau III.17.

$b_1(\text{nm})$	$b_2(\text{nm})$	$h(\text{nm})$	$h_r(\text{nm})$
$29,6 \pm 2,1$	$291 \pm 4,9$	$221,9 \pm 4,2$	$34,1 \pm 3,1$

TAB. III.17: Paramètres géométriques du réseau E_{hr}^5 estimés par le PMC dans le cas d'un profil trapézoïdal avec couche résiduelle.

Ce résultat constitue une seconde vérification de la présence de la couche résiduelle au fond du réseau. La valeur de l'épaisseur h_r a été également calculée par la loi de conservation de la matière (équation III.9) et vaut dans ce cas 35 nm, ce qui est cohérent avec la valeur estimée par le RN.

Cette partie a fait l'objet d'une étude de la classification neuronale dans le domaine de la scatterométrie. Nous avons montré qu'un RN classifieur permet de prédire le modèle géométrique d'un réseau de diffraction à partir de sa seule signature ellipsométrique mesurée. La première application nous a montré que le RN est ainsi capable de dissocier deux profils géométriques ayant deux signatures ellipsométriques proches. La seconde application, qui est toujours à l'étude, nous a permis de concevoir un classifieur pouvant détecter un défaut de fabrication nuisible à la lithographie, à savoir une couche résiduelle au fond du masque de résine. L'épaisseur de cette couche doit être supérieure à environ 20 nm pour que le taux d'erreur soit faible (5%) avec une signature optique employée.

III.6 Conclusion

Dans ce chapitre, nous avons présenté la méthode neuronale appliquée à la scatterométrie pour la reconstruction du profil géométrique des réseaux de diffraction. Le principe

est basé sur l'emploi d'un réseau de neurones dans la phase de la résolution du problème inverse. Ce traitement s'effectue dans un domaine précis de l'espace des paramètres géométriques défini préalablement par l'utilisateur lors de la phase d'apprentissage. Une base de données constituée d'un grand nombre de couples {signatures ellipsométriques / paramètres géométrique} doit être simulée pour permettre au RN d'apprendre la relation qui existe entre la signature ellipsométrique et les paramètres géométriques recherchés.

Nous avons montré la capacité du RN à caractériser des réseaux de diffraction possédant diverses formes de profils géométriques. Les résultats ont été vérifiés et comparés lorsque cela était possible avec des caractérisations microscopiques. En scatterométrie, la période du réseau est souvent considérée une hypothèse sous jacente et ainsi supposée parfaitement déterminée soit directement par la technique de fabrication soit par une mesure alternative. Or, dans certaines conditions ce n'est pas le cas. Ainsi, nous avons montré que le réseau de neurones pouvait facilement accéder à ce paramètre et le déterminer avec précision.

L'application de la méthode de sélection HVS nous a permis de réduire considérablement le temps de caractérisation globale. D'une part, elle supprime les mesures non pertinentes qui ne portent pas d'informations supplémentaires utiles pour la caractérisation. D'autre part, elle simplifie l'architecture du RN en choisissant un nombre de neurones cachés optimal tout en gardant des mêmes performances optimales. Il devient ainsi aisé de contrôler de façon très rapide l'homogénéité des échantillons.

Nous avons mené une étude paramétrique dans le cadre théorique, afin d'examiner l'influence de certains paramètres sur le comportement du RN, plus particulièrement les hypothèses du problème qui sont souvent fixées par l'utilisateur. Il est ainsi primordial de connaître certains d'entre eux avec précision, comme la période lorsqu'elle est supposée fixe et l'angle d'azimut. Nous avons montré également l'influence d'une hypothèse fautive du modèle géométrique, fixé lors de l'entraînement du RN. Ce dernier donne la plupart des cas un résultat valide (contenu dans son domaine d'entraînement) mais qui s'écarte du profil géométrique réel qu'on est censé mesurer. La seule façon pour valider nos résultats reste le calcul d'une erreur quadratique entre la signature calculée et celle mesurée sur des données non biaisées, si possible pour un grand nombre de configurations expérimentales autre que celles utilisées pour l'entraînement.

Finalement, nous avons proposé une autre application des RN qui peut être utile dans le cadre précis de la scatterométrie : la classification. En effet, nous avons employé cette fonctionnalité dans le but d'identifier le modèle géométrique du réseau avant la phase de caractérisation. Dans la première application étudiée, nous avons montré que le RN était capable de classer les signatures selon deux classes correspondantes à deux profils géométriques distincts, à savoir le profil trapézoïdal simple et le profil sinusoïdal. La deuxième application proposée vise à détecter un défaut particulier lors du dépôt des masques en résine, à savoir la présence d'une couche résiduelle au fond du réseau. Le classifieur mis en place permet d'attribuer à la signature un profil trapézoïdal simple ou avec couche résiduelle, avec un certain taux d'erreur fixé par l'utilisateur.

Conclusion et perspectives

Dans ce manuscrit, nous avons présenté une méthode de caractérisation optique qui peut constituer une véritable alternative aux méthodes classiques. Elle est basée sur le couplage d'une mesure ellipsométrique avec un traitement par réseaux de neurones pour la résolution du problème inverse. Les contributions des réseaux de neurones en métrologie optique et dimensionnelle ont été étudiées au cours de cette thèse. Nous avons concentré nos efforts sur la caractérisation des structures en couches minces ainsi que les réseaux de diffraction. De part ses capacités d'apprentissage, le réseau de neurones peut ainsi apprendre la relation qui lie la signature optique mesurée sur un échantillon à ses paramètres optiques et géométriques le caractérisant. Pour ce faire, une base de données doit être préalablement construite. Elle est constituée d'un grand nombre de signatures optiques simulées à partir d'un modèle structural donné, décrit par un certain nombre de paramètres, appartenant à un domaine bien défini de l'espace, fixé par l'utilisateur. Les études se poursuivront afin de concevoir des réseaux de neurones fonctionnant sur des domaines plus larges de l'espace des paramètres. Nous avons apporté une première solution par l'utilisation d'un classifieur travaillant en amont. Cette voie continue d'être exploitée. La durée de calcul de la base dépend fortement du type de structure étudiée. Il est par exemple beaucoup moins long pour les structures en multicouches que pour les structures diffractantes. Le réseau va être ensuite entraîné sur ces exemples afin d'approcher au mieux la relation souhaitée. Bien que la durée de l'entraînement nécessite quelques minutes, la phase d'exploitation est instantanée. Cette propriété intéressante du réseau de neurones peut être mise à profit dans le cadre d'une caractérisation en temps réel, dans un processus de fabrication par exemple. La méthode de sélection HVS employée contribue dans le même sens à rendre la caractérisation neuronale plus rapide, en diminuant la taille de la signature optique en supprimant les points de mesure non porteurs d'informations, et en réduisant l'architecture intrinsèque du réseau par un choix optimal de neurones dans sa couche cachée. La durée nécessaire à mesurer la signature optique, et à entraîner le réseau se trouve ainsi considérablement réduite facilitant un contrôle encore plus rapide dans une chaîne de fabrication.

Dans le cadre précis de la caractérisation de couches minces, nous avons montré que le réseau de neurones est un outil efficace pour la détermination des paramètres optiques et

géométriques comme l'indice de réfraction et l'épaisseur. Sa relative résistance aux minima locaux le rend performant comparé aux méthodes classiques d'optimisation comme par exemple l'algorithme de Levenberg-Marquardt. Il peut également être adapté à d'autres types de structures plus complexes possédant par exemple plusieurs couches ou en ajoutant simplement une couche de rugosité. C'est dans cette voie que les travaux vont se poursuivre. Nous avons montré également l'avantage de la classification neuronale pour la métrologie sur un exemple simple. En effet, nous avons proposé la détermination de la gamme d'épaisseur de couches minces à partir de signatures optiques. Cette application permet ensuite d'aiguiller vers le réseau de neurones le plus efficace. Elle peut également être couplée aux méthodes d'optimisation classiques. En effet, la localisation de la solution dans un domaine restreint de l'espace des paramètres permet de choisir efficacement le point de départ nécessaire pour ces méthodes. Ceci permet de même d'éviter certains minima locaux. D'autres applications sont envisageables dans ce domaine et constitue également la suite à donner à ce travail, citons notamment la détection de modèle structurale (couche supplémentaire, rugosité, etc.).

Nous avons également montré différentes contributions du réseau de neurones en scatterométrie. Ainsi, nous avons démontré la flexibilité de la méthode par la caractérisation de structures possédant des profils géométriques divers. Nous avons développé par ailleurs une méthode qui permet de déterminer la période avec précision. Habituellement, cette détermination est réalisée en utilisant le montage de Littrow, mais ce dernier devient vite limité dans le cas des réseaux de faible période où aucun ordre de diffraction supérieur ne se propage. De plus l'utilisation de la classification neuronale nous a permis une nouvelle application en scatterométrie, à savoir l'identification des modèles géométriques avant la phase de traitement. Cette application peut présenter un intérêt non négligeable en lithographie par la détection automatique de la présence d'une couche résiduelle indésirable au fond du masque de résine nuisible pour l'étape de gravure ultérieure. Cette étude est toujours en cours de développement au sein du laboratoire pour améliorer la limite détectable de l'épaisseur résiduelle. A cet effet, on s'intéressera à la sensibilité sur la signature ellipsométrique de la couche.

D'après toutes les applications énoncées, une procédure de caractérisation entièrement neuronale est envisageable et en cours de préparation et de développement. En effet, à partir de la signature optique mesurée, il est prévu d'employer en cascade toutes les applications du réseau de neurones (classification, sélection, caractérisation) afin d'offrir une caractérisation neuronale plus performante en terme de rapidité et d'efficacité.

Ce travail de recherche a ouvert la voie à de nouvelles applications envisageables dans le futur dans le cadre de la caractérisation optique. Le réseau de neurones peut ainsi être adapté à la caractérisation de plusieurs types de structures. A titre d'exemple, il est fortement souhaité dans le cas des réseaux 2D, de plus en plus employés en pratique pour des applications optiques. En effet, sa rapidité de traitement constitue un réel contraste avec les méthodes classiques qui utilise le problème direct relativement long dans ce cas. La rugosité des flancs d'un réseau de diffraction est également un paramètre nuisible puisqu'il modifie ses propriétés notamment lorsqu'il est fabriqué à très basse échelle. Le besoin actuel de caractériser cette rugosité nous incitera dans un avenir proche à envisager une solution dans le cadre de notre méthode.

Les résultats satisfaisants obtenus au cours de cette thèse nous encourage à continuer d'exploiter les réseaux de neurones dans des domaines connexes, comme l'optique guidée, où ils ont déjà montré leur intérêt [180].

Annexe A : Indices optiques des différents matériaux utilisés

Dans cet annexe, nous donnerons les indices optiques des différents matériaux étudiés dans la gamme des longueurs d'onde exploitées dans ce travail.

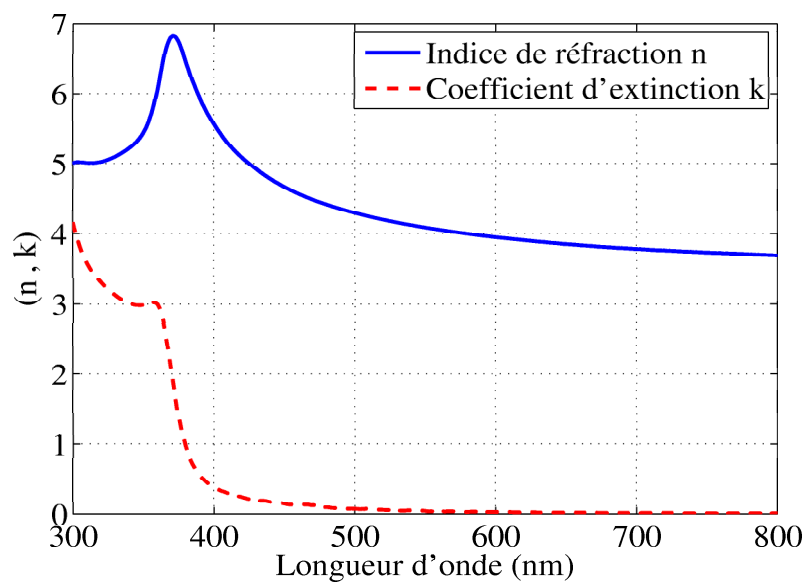


FIG. III.53: Indice de réfraction n et coefficient d'extinction k du silicium.

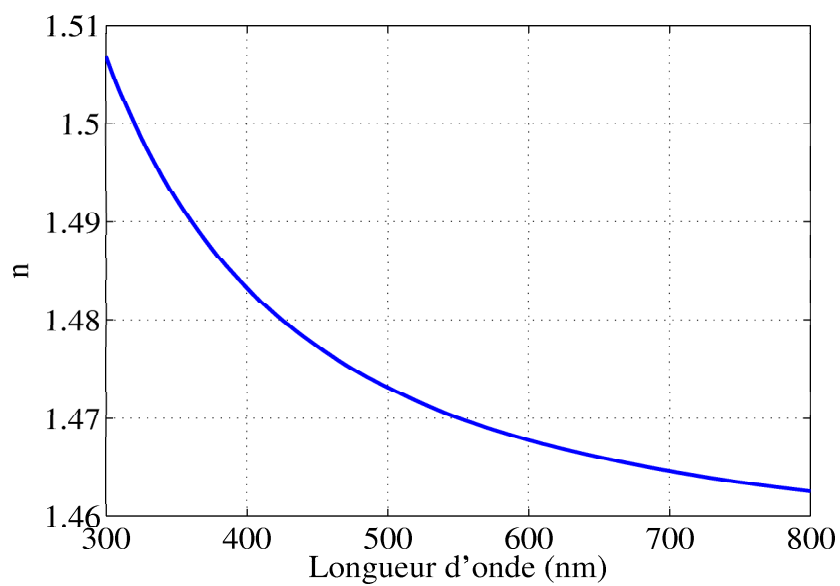


FIG. III.54: Indice de réfraction n de la silice.

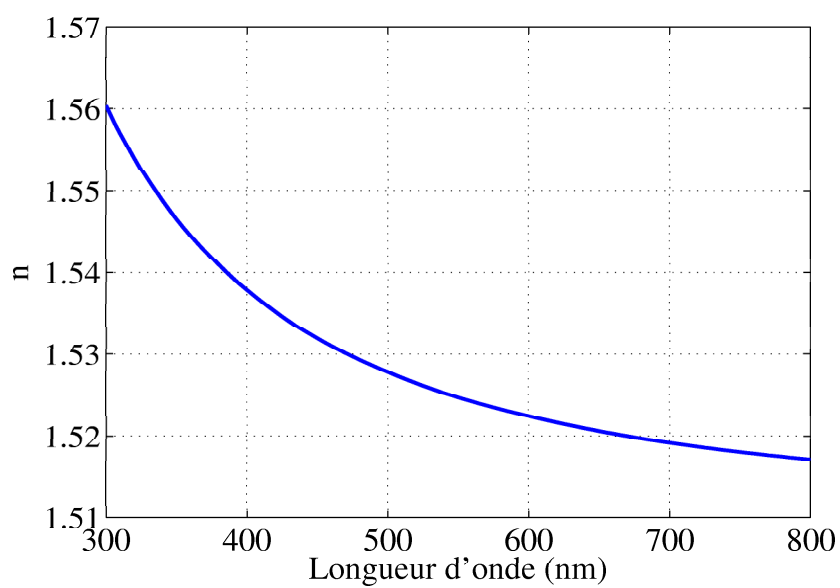


FIG. III.55: Indice de réfraction n du verre BK_7 .

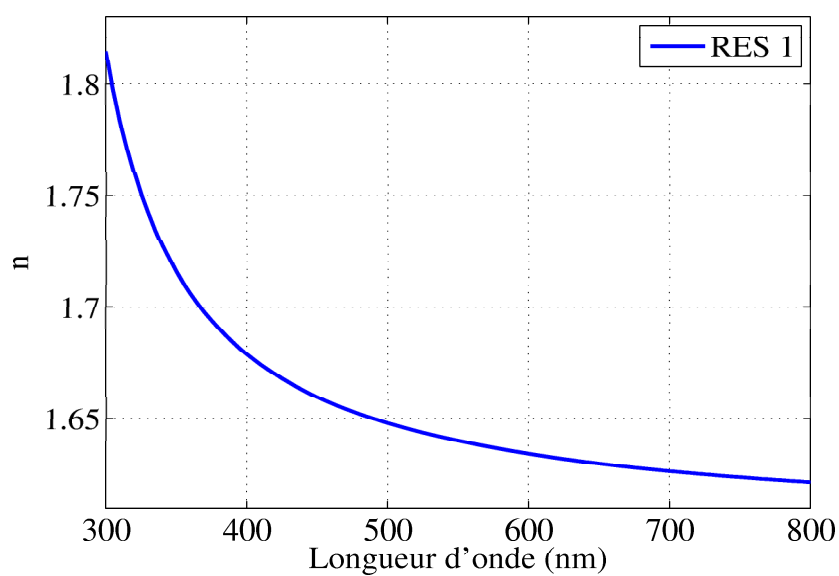


FIG. III.56: Indice de réfraction n de la résine utilisée par le LHC, noté RES1.

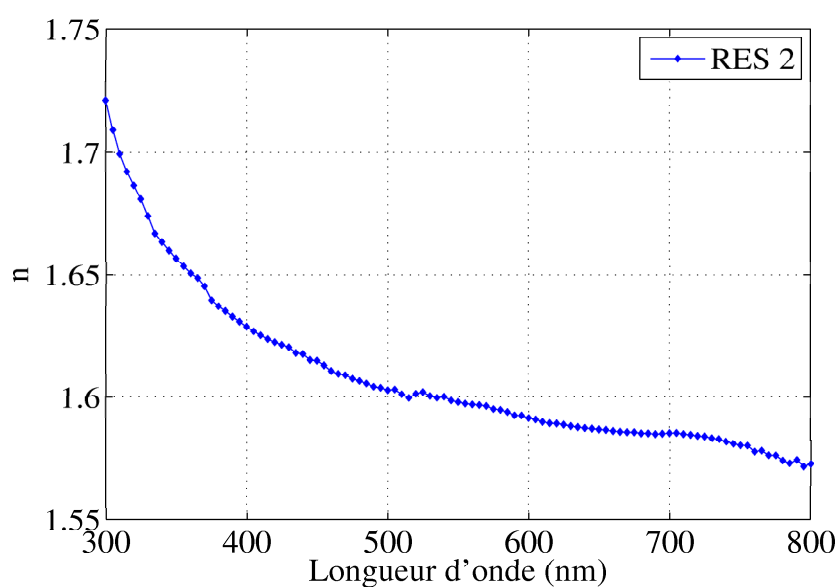


FIG. III.57: Indice de réfraction n de la résine utilisée par le LTM, noté RES 2, calculé point par point dans la gamme des longueurs d'onde étudiée.

Annexe B : Algorithme de rétropropagation de l'erreur

Dans le présent annexe, on détaillera le principe de l'algorithme de rétropropagation du gradient de l'erreur. On rappelle que ce dernier n'est pas un algorithme d'apprentissage, il est plutôt utilisé dans une procédure d'apprentissage pour d'une manière récursive évaluer le gradient de la fonction coût.

Considérons un réseau de neurones de type PMC, avec comme paramètres :

- N_c et N_s : nombre de neurones dans la couche cachée et la couche de sortie, respectivement,
- u_j et v_m : potentiel du neurone caché j et du neurone de sortie m , respectivement,
- f et g : fonctions de transfert des neurones de la couche cachée et la couche de sortie, respectivement. Pour simplifier le problème, on choisit f sigmoïdale et g linéaire,
- W et Z : matrices contenant les poids synaptiques entre le vecteur d'entrée et la couche cachée (dimension $[N_c \times N]$), et la couche cachée et la couche de sortie (dimension $[N_s \times N_c]$), respectivement,
- μ : exemple quelconque dans l'ensemble des données d'entraînement ($\mu : 1, \dots, n$),
- x_i^μ : composante i du vecteur d'entrée correspondant à l'exemple μ ($i : 1, \dots, N$).

Rappelons que le potentiel d'un neurone est une somme pondérée de ses entrées, on a ainsi, pour un exemple (ou vecteur d'entrée) μ :

$$u_j^\mu = \sum_{i=0}^N w_{j,i} x_i^\mu \quad (\text{III.18})$$

$$v_j^\mu = \sum_{j=0}^{N_c} z_{m,j} f(u_j^\mu) \quad (\text{III.19})$$

Pour évaluer le gradient de la fonction coût $C(W, Z)$ sur l'ensemble des exemples d'entraînement, il suffit de sommer les gradients calculés sur chaque exemple $C^\mu(W, Z)$:

$$C(W, Z) = \sum_{\mu=1}^n C^\mu(W, Z) \quad (\text{III.20})$$

avec :

$$C^\mu = \sum_{m=1}^{N_s} (y_m^\mu - s_m^\mu)^2 \quad (\text{III.21})$$

La sortie du réseau s_m^μ calculée pour l'exemple μ est donnée par :

$$s_m^\mu = g \left(\sum_{j=0}^{N_c} z_{m,j} f(u_j^\mu) \right) = g(v_m) = v_m^\mu \quad (\text{III.22})$$

En premier temps, l'influence des paramètres Z sur les sorties du réseau est évaluée par le calcul de la dérivée de la fonction coût par rapport à ces derniers :

$$\left(\frac{\partial C^\mu}{\partial z_{m,j}} \right) = \left(\frac{\partial C^\mu}{\partial v_m} \right) \left(\frac{\partial v_m}{\partial z_{m,j}} \right) = \delta_m^\mu f(u_j^\mu) \quad (\text{III.23})$$

La quantité δ_m^μ s'exprime simplement, à partir des équations III.22 et III.20, par :

$$\begin{aligned} \delta_m^\mu &= \left(\frac{\partial C^\mu}{\partial v_m} \right) = \frac{\partial}{\partial v_m} \left(\sum_{k=1}^{N_s} (y_k^\mu - s_k^\mu)^2 \right) \\ &= -2 \sum_{k=1}^{N_s} \left[(y_k^\mu - s_k^\mu) \left(\frac{\partial s_k^\mu}{\partial v_m} \right) \right] \\ &= -2 (y_k^\mu - s_k^\mu) \end{aligned} \quad (\text{III.24})$$

ce qui donne :

$$\left(\frac{\partial C^\mu}{\partial z_{m,j}} \right) = -2 f(u_j^\mu) (y_k^\mu - s_k^\mu)^2 \quad (\text{III.25})$$

Pour évaluer l'influence des paramètres W , on procède de la façon suivante :

$$\begin{aligned} \left(\frac{\partial C^\mu}{\partial w_{j,i}} \right) &= \left(\frac{\partial C^\mu}{\partial u_j} \right) \left(\frac{\partial u_j}{\partial w_{j,i}} \right) \\ &= \left[\sum_{m=1}^{N_s} \left(\frac{\partial C^\mu}{\partial v_m} \right) \left(\frac{\partial v_m}{\partial u_j} \right) \right] x_i^\mu \\ &= \left[\sum_{m=1}^{N_s} \delta_m^\mu \left(\frac{\partial v_m}{\partial u_j} \right) \right] x_i^\mu \end{aligned} \quad (\text{III.26})$$

or d'après l'équation III.18, on a :

$$\left(\frac{\partial v_m}{\partial u_j} \right) = z_{m,j} \left(\frac{\partial f}{\partial u_j} \right) \quad (\text{III.27})$$

ce qui donnera à la fin :

$$\begin{aligned} \left(\frac{\partial C^\mu}{\partial w_{j,i}} \right) &= \left[\sum_{m=1}^{N_s} \delta_m^\mu z_{m,j} \left(\frac{\partial f}{\partial u_j} \right) \right] x_i^\mu \\ &= x_i^\mu \left(\frac{\partial f}{\partial u_j} \right) \sum_{m=1}^{N_s} \delta_m^\mu z_{m,j} \end{aligned} \quad (\text{III.28})$$

En résumé, l'algorithme de rétropropagation calcule d'abord les sorties du réseau de

neurones pour les exemples utilisés pour l'entraînement, à l'aide des entrées correspondantes et les poids W et Z , c'est la phase de propagation. L'erreur sur le neurone caché j est calculée en sommant les gradients sur les neurones de sorties pondérés par l'influence de la connexion qui les relie au neurone j par le poids $z_{m,j}$.

Bibliographie

- [1] H. Sahseh. *Conception et réalisation d'un nouvel ellipsomètre automatisé, achromatique, à modulation de polarisation Faraday*. PhD thesis, Université de Saint Etienne, 1994.
- [2] P. Drude. *Ann. Phys.*, 32 :584, 1887.
- [3] P. Drude. *Ann. Phys.*, 34 :489, 1889.
- [4] P. Drude. *Ann. Phys.*, 36 :532, 1889.
- [5] A. Rothen. The ellipsometer, an apparatus to measure thicknesses of thin surface films. *Rev. Sci. instrum.*, 16 :26–30, 1945.
- [6] P. Boher, P. Evrard, J. P. Piel, C. Defranoux, and J. L. Stehle. Characterizing resists and films with vuv spectroscopic ellipsometry. *Sol. State Tech.*, 1 :pp 165–175, 2001.
- [7] S. Marceau, J. H. Tortai, J. Tillier, N. Vourdas, E. Gogolides, I. Raptis, K. Beltsios, and K. van Werden. Thickness-dependent glass transition temperature of thin resist films for high resolution lithography. *Microelec. Eng.*, 83 :pp 1073–1077, 2006.
- [8] M. Mansour, A. En Naciri, L. Johann, S. Duguay, J. J. Grob, M. Stchakovsky, and C. Eypert. Dielectric function of germanium nanocrystals between 0.6 and 6.5 ev by spectroscopic ellipsometry. *J. Phys. Chem. Sol.*, 67 :1291–1294, 2006.
- [9] M. Gaillet, L. Yan, and E. Teboul. Optical characterizations of complete tft-lcd display devices by phase modulated spectroscopic ellipsometry. *Th. Sol. Fil.*, 516 :pp 170–174, 2007.
- [10] D. Bhattacharyya, N. K. Sahoo, S. Thakur, and N. C. Das. Characterization of a multilayer highly reflecting mirror by spectroscopic ellipsometry. *App. Opt.*, 40 :1707–1714, 2001.
- [11] E. S. Kooij, A. C. Galca, and B. Poelsema. Versatile transmission ellipsometry to study linear ferrofluid magneto-optics. *J. Coll. Int. Sc.*, 304 :pp 261–270, 2006.

- [12] F. Choueikani. *Etude des potentialités de couches minces sol-gel dopées par des nanoparticules magnétiques pour la réalisation de composants magnéto-optiques intégrés*. PhD thesis, Université Jean Monnet de Saint Etienne, 2008.
- [13] D. K. Goyal, G. K. Pribil, J. A. Woollam, and A. Subramanian. Detection of ultrathin biological films using vacuum ultraviolet spectroscopic ellipsometry. *Mat. Sc. Eng. B*, 149 :pp 26–33, 2008.
- [14] E. Garcia-Caurel, B. Dré villon, A. De Martino, and L. Schwartz. Application of fourier transform infrared ellipsometry to assess the concentration of biological molecules. *App. Opt.*, 41 :pp 7339–7345, 2002.
- [15] J. Lee and R. Collins. Real-time characterization of film growth on transparent substrates by rotating-compensator multichannel ellipsometry. *App. Opt.*, 37 :pp 4230–4238, 1998.
- [16] M. G. Boudreau, S. G. Wallace, G. Balcaitis, S. Murukgar, H. K. Haugen, and P. Mascher. Application of in situ ellipsometry in the fabrication of thin-film optical coatings on semiconductors. *App. Opt.*, 39 :pp 1053–1058, 2000.
- [17] D. E. Aspnes. Expanding horizons : new developments in ellipsometry and polarimetry. *Th. Sol. Fil.*, 455-456 :3–13, 2004.
- [18] R. M. A. Azzam and N. M. Bashara. *Ellipsometry and polarized light*. 2nd ed., 1987.
- [19] H. G. Tompkins. *A user's guide to ellipsometry*. Academic Press, 1993.
- [20] D. E. Aspnes and A. A. Studna. High precision scanning ellipsometer. *App. Opt.*, 14 :pp 220–228, 1975.
- [21] Y. J. van der Meulen and N. C. Hien. Design and operation of an automated, high-temperature ellipsometer. *J. Opt. Soc. Am. A*, 64 :pp 804–811, 1974.
- [22] J. N. M. de Nijs and A. van Sifflhout. Systematic and random errors in rotating-analyzer ellipsometry. *J. Opt. Soc. Am. A*, 5 :773–781, 1988.
- [23] P. Chindaudom and K. Vedam. Determination of the optical function $n(\lambda)$ of vitreous silica by spectroscopic ellipsometry with an achromatic compensator. *App. Opt.*, 32 :pp 6391–6398, 1993.
- [24] F. Bertin, A. Chabli, E. Chiariglione, M. Burdin, M. Berger, T. Boudet, O. Lartigue, and G. Ravel. Spectroscopic ellipsometry with compensator and x-ray specular reflectivity for characterization of thin optical layers on transparent substrates. *Th. Sol. Fil.*, 313-314 :pp 68–72, 1998.
- [25] J. Opsal, J. Fanton, J. Chen, J. Leng, L. Wei, C. uhrich, M. Senko, C. Zaiser, and D. E. Aspnes. Broadband spectral operation of a rotating-compensator ellipsometer. *Th. Sol. Fil.*, 313-314 :pp 58–61, 1998.
- [26] R. Kleim, L. Kuntzler, and A. El Ghemmaz. Systematic errors in rotating-compensator ellipsometry. *J. Opt. Soc. Am. A*, 11 :p 2550–2559, 1994.

- [27] D. E. Aspnes. Optimizing precision of rotating-analyzer and rotating-compensator ellipsometers. *J. Opt. Soc. Am. A*, 21 :pp 403–410, 2004.
- [28] L. Broch, A. EN Naciri, and L. Johann. Accurate calibration and optimized measurement : use of an achromatic compensator in rotating polarizer spectroscopic ellipsometry. *Th. Sol. Fil.*, 455-456 :pp 61–65, 2004.
- [29] S. N. Jasperson and S. E. Schnatterly. An improved method for high reflectivity ellipsometry based on a new polarization modulation technique. *Rev. Sci. Instrum.*, 40 :pp 761–767, 1969.
- [30] O. Acher, E. Bigan, and B. Dré villon. Improvements of phase-modulated ellipsometry. *Rev. Sci. Instrum.*, 60 :pp 65–77, 1989.
- [31] M. W. Wang, F. H. Tsai, and Y. F. Chao. In situ calibration technique for photoelastic modulator in ellipsometry. *Th. Sol. Fil.*, 455-456 :PP 78–83, 2004.
- [32] K. Yoshioka, T. Sameshima, and N. Sano. Progress in fabrication processing of thin film transistors. *Solide-State Electr.*, 52 :359–364, 2008.
- [33] S. Uchikoga and N. Ibaraki. Low temperature poly-si tft-lcd by excimer laser anneal. *Th. Sol. Fil.*, 19, 2001.
- [34] B. K. Crone, A. Dodabalapur, R. Sarpeshkar, A. Gelperin, H. Katz, and Z. Bao. Organic oscillator and adaptive amplifiers for chemical vapor sensing. *J. Appl. Phys.*, 91 :10140–10146, 2002.
- [35] B. Kaplan, T. Novikova, A. De Martino, and B. Dré villon. Characterization of bi-dimensional gratings by spectroscopic ellipsometry and angle-resolved mueller polarimetry. *App. Opt.*, 43 :1233–1240, 2004.
- [36] C. R. A. Lima, L. L. Soares, L. Cescato, and A. L. Gobbi. Reflecting polarizing beam splitter. *Opt. Lett.*, 22 :203–205, 1997.
- [37] P. Lalanne, J. Hazart, P. Chavel, E. Cambril, and H. Launois. A transmission polarizing beam splitter grating. *J. Opt. A, Pure Appl. Opt.*, 1 :215, 1999.
- [38] B. Wang, C. Zhou, S. Wang, and J. Feng. Polarizing beam splitter of a deep-etched fused-silica grating. *Opt. Lett.*, 32 :1299–1301, 2007.
- [39] R. Waldhausl, B. Schnabel, P. Dannberg, E. B. Kley, A. Bra uer, and W. Karthe. Efficient coupling into polymer waveguides by gratings. *App. Opt.*, 36 :9383–9390, 1997.
- [40] J. Vorosa, J.J. Ramsdenb, G. Csucsc, I. Szendrod, S.M. De Paula, M. Textora, and N.D. Spencera. Optical grating coupler biosensors. *Biomatériaux*, 23 :3699–3710, 2002.
- [41] A.-L. Ferembach. *Réseaux résonnants à bande interdite photonique, nouveaux filtres pour la D.W.D.M.* PhD thesis, université Aix-Marseille III, Marseille, 2003.

- [42] J. M. Ostermann, P. Debernardi, C. Jalics, A. Kroner, M. C. Riedl, and R. Michalzik. Surface gratings for polarization control of single and multi-mode oxide-confined vertical-cavity surface-emitting lasers. *Opt. Comm.*, 246 :511–519, 2005.
- [43] M. Niggemann, M. Glatthaar, A. Gombert, A. Hinsch, and V. Wittwer. Diffraction gratings and buried nano-electrodes—architectures for organic solar cells. *Th. Sol. Fil.*, 451-452 :619–623, 2004.
- [44] F. El-Daisty and A. Amichi. Evanescent-wave in perturbed optical fibers. *Opt. Com.*, 281 :4329–4333, 2008.
- [45] R. W. Wood. On a remarkable case of uneven distribution of light in a diffraction grating spectrum. *Philos. Mag.*, 4 :396–402, 1902.
- [46] E. G. Loewen, M. Nevière, and D. Maystre. On an asymptotic theory of diffraction gratings used in the scalar domain. *J. Opt. Soc. Am. A*, 68 :496–502, 1978.
- [47] D. Maystre. Sur la diffraction d’une onde plane par un réseau métallique de conductivité finie. *Opt. Comm.*, 6 :pp 50–54, 1972.
- [48] R. Petit. Etude théorique de la diffraction d’une onde plane et monochromatique par un réseau métallique infiniment conducteur. *App. Opt.*, 4 :pp 1551–1554, 1965.
- [49] R. Petit. Contribution à l’étude de la diffraction par un réseau métallique. *Rev. Opt.*, 45 :pp 249–276, 1966.
- [50] T. Delort and D. Maystre. Finite-element method for gratings. *J. Opt. Soc. Am. A*, 10 :2592–2601, 1993.
- [51] G. Tayeb. The method of fictitious sources applied to diffraction gratings. *Appl. Comput. Electromag. Soc.*, 9 :90–100, 1994.
- [52] A. V. Tishchenko. A generalized source method for wave propagation. *Pure and App. Opt. A*, 7 :1425–1449, 1998.
- [53] A. V. Tishchenko. Generalized source method : new possibilities for waveguide and grating problems. *Opt. and Quant. Elec.*, 32 :971–980, 2000.
- [54] M. Nevière, G. Cerutti-Maori, and M. Cadilhac. Sur une nouvelle méthode de résolution du problème de la diffraction d’une onde plane par un réseau infiniment conducteur. *Opt. Comm.*, 3 :pp 48–52, 1971.
- [55] R. Petit. *Electromagnetic theory of gratings*. Springer-Verlag, Berlin, 1980.
- [56] L. Li. Use of fourier series in the analysis of the discontinuous structures. *J. Opt. Soc. Am. A*, 13 :pp 1870–1876, 1996.
- [57] M. Nevière and E. Popov. *Light Propagation in Periodic Media : Differential Theory and Design*. Marcel Dekker, 2003.
- [58] L. Li. Multilayer modal method for diffraction gratings of arbitrary profile, depth, and permittivity. *J. Opt. Soc. Am. A*, 10 :pp 2583–2591, 1993.

- [59] J. Chandezon, D. Maystre, and G. Raoult. A new theoretical method for diffraction gratings and its numerical application. *J. Optics (Paris)*, 11 :pp 235–241, 1980.
- [60] G. Granet and J. Chandezon. The method of curvilinear coordinates applied to the problem of scattering from surface-relief gratings defined by parametric equations : application to scattering from a cycloidal grating. *Pure Appl. Opt.*, 6 :727–740, 1997.
- [61] J. R. Anderwartha, G. H. Derrick, and R. C. McPhedran. A general modal theory for reflection gratings. *Opt. Acta*, 28 :pp 1501–1516, 1981.
- [62] P. Sheng, R. S. Stepleman, and P. N. Sandra. Exact eigenfunctions for square-wave gratings : Application to diffraction and surface-plasmon calculations. *Phys. Rev. B*, 26 :pp 2907–2916, 1982.
- [63] M. G. Moharam and T. K. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *J. Opt. Soc. Am. A*, 71 :pp 811–818, 1981.
- [64] M. G. Moharam and T. K. Gaylord. Diffraction analysis of dielectric-surface relief gratings. *J. Opt. Soc. Am. A*, 72 :1385–1392, 1982.
- [65] M. G. Moharam and T. K. Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction - e-mode polarization and losses. *J. Opt. Soc. Am. A*, 73 :451–455, 1983.
- [66] M. G. Moharam and T. K. Gaylord. Rigorous coupled wave analysis of metallic surface-relief gratings. *J. Opt. Soc. Am. A*, 3 :1780–1796, 1986.
- [67] M. G. Moharam and T. K. Gaylord. Three dimensional vector-coupled wave analysis of planar grating diffraction. *J. Opt. Soc. Am. A*, 73 :pp 1105–1112, 1983.
- [68] M. Moharam. Coupled-wave analysis of two-dimensional dielectric gratings. *Proc. SPIE*, 883 :8–11, 1988.
- [69] K. Rokushima and J. Yamakita. Analysis of anisotropic dielectric gratings. *J. Opt. Soc. Am. A*, 73 :901–908, 1983.
- [70] L. Li. Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings. *J. Opt. Soc. Am. A*, 13 :pp 1024–1035, 1996.
- [71] P. Lalanne and G. M. Morris. Highly improved convergence of the coupled wave method for tm polarization. *J. Opt. Soc. Am. A*, 13 :pp 779–784, 1996.
- [72] G. Granet and B. Guizal. Really efficient implementation of the coupled-wave method for metallic lamellar gratings in tm polarization. *J. Opt. Soc. Am. A*, 13 :pp 1019–1023, 1996.
- [73] H. Curry. The method of steepest descent for nonlinear minimization problems. *Quart. Appl. Math.*, 2 :258–261, 1944.
- [74] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *Computer J.*, 7 :149–154, 1964.

- [75] E. Polak. *Computational methods in optimization, a unified approach*. Academic Press, New York, 1971.
- [76] G. C. Broyden. A class of method for solving nonlinear simultaneous equations. *Math. Comput.*, pages 577–593, 1965.
- [77] R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *Comp. J.*, 6 :163–168, 1963.
- [78] D. E. Goldberg. Conjugate gradient method for non linear programming problems with linear constraints. *I & EC Fundam.*, pages 142–151, 1968.
- [79] D. F. Shanno. An accelerated gradient projection method for linearly constrained nonlinear estimation. *SIAM J. Appl. Math.*, pages 322–334, 1970.
- [80] K. Levenberg. A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, 2 :pp 164–168, 1944.
- [81] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 13 :pp 431–441, 1963.
- [82] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7 :308–313, 1965.
- [83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220 :671–680, 1983.
- [84] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, and A.H. Teller. Equations of state calculations by fast computing machines. *J. of Chem. Phys.*, 21 :1087–1091, 1953.
- [85] O. Polgar, M. Fried, T. Lohner, and I. Barsony. Comparison of algorithms used for evaluation of ellipsometric measurements random search, genetic algorithms, simulated annealing and hill climbing graph-searches. *Surf. Sci.*, 457 :157–177, 2000.
- [86] Y. Zhaoxian and M. Dang. Generalized simulated annealing algorithm applied in the ellipsometric inversion problem. *Th. Sol. Fil.*, 425 :108–112, 2003.
- [87] J. H. Holland. *Adaptation in natural and artificial systems*. 1975.
- [88] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [89] G. Cormier and R. Boudreau. Genetic algorithm for ellipsometric data inversion of absorbing layers. *J. Opt. Soc. Am. A*, 17 :129–134, 2000.
- [90] S. Robert. *Caractérisation optique de réseaux de diffraction submicroniques par des techniques génétique et neuronale*. PhD thesis, Université Jean Monnet, 2003.
- [91] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge university press, 1992.
- [92] D. Hebb. *The Organization of Behavior*. Wiley : New York, 1949.

- [93] F. Rosenblatt. The perceptron : a perceiving and recognizing automaton. Technical report, Cornell Aeronautical Lab, 1957.
- [94] M. L. Minsky and S. A. Papert. *Perceptrons*. Cambridge, MIT Press, 1969.
- [95] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1 :541–551, 1989.
- [96] I. Rivals. *Modélisation et commande de processus par réseaux de neurones : application au pilotage d'un véhicule autonome*. PhD thesis, Université Pierre et Marie Curie de Paris, 1995.
- [97] H. Gonzalez-Diaza, I. Bonetb, C. Terand, E. De Clercq, R. Bellob, M. Garciab, L. Santanaa, and E. Uriarte. Ann-qsar model for selection of anticancer leads from structurally heterogeneous series of compounds. *European Journal of Medicinal Chemistry*, 42 :580–585, 2007.
- [98] J. Brajard, C. Jamet, C. Moulin, and S. Thiria. Use of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean colour sensor : Application to absorbing aerosols. *Neur. Net.*, 19 :178–185, 2006.
- [99] A. Aussem. *Théorie et applications des réseaux de neurones récurrents et dynamiques à la prédiction, à la modélisation et au contrôle adaptatif des processus dynamiques*. PhD thesis, Universit de Paris 05, Paris, 1995.
- [100] G. Cybenko. Approximation by superpositions of sigmoïdal function. *Mathematics of Control, Signals and Systems*, 2 :303–314, 1989.
- [101] K. Funahashi. On the approximate relaization of continuous mapping by neural networks. *Neural Networks*, 2 :183–192, 1989.
- [102] J. Zhang, G. G. Walter, Y. Miao, and N. Lee Wayne. Wavelet neural networks for function learning. *IEEE Trans. Signal Process.*, 6 :1485–1497, 1995.
- [103] K. Hornik, M. Stinchcombe, H. White, and P. Auer6. Degree of approximation results for feedforward networks approximating unknown mapping and their derivatives. *Neural Computation*, 6 :1262–1275, 1994.
- [104] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error back-propagation. *Parallel Distributed Processing : Explorations in the Microstructure of Cognitions*, pages 318–362, 1986.
- [105] M. McInerney and A.P. Dhawan. Use of genetic algorithms with backpropagation in training offeedforward neural networks. In *IEEE International Conference on Neural Networks*, 1993.
- [106] V. Maniezzo. Genetic evolution of the topology and weight distribution of neural-networks. *IEEE Transactions on Neural Networks*, 5 :39–53, 1994.
- [107] D. Nguyen and B. Widrow. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proc. Int. Join. Conf. on Neur. Networks*, 3 :21–26, 1990.

- [108] I. T. Jolliffe. *Principal component analysis*. Springer, 1986.
- [109] P. Demartines. *Analyse de données par réseaux de neurones auto-organisés*. PhD thesis, Institut national polytechnique de Grenoble, 1995.
- [110] J. Hérault and C. Jutten. *Réseaux de neurones et traitement du signal*. Hermès, 1993.
- [111] T. Kohonen. *Self organizing maps*. Springer, 3^e édition, 2001.
- [112] S. German, A. Bienenstock, and R. Doursat. Neural networks and the biais/variance dilemma. *Neural Computation*, 4 :1–58, 1992.
- [113] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, 8 :349–359, 1974.
- [114] S. Saarinen, R. Bramley, and G. Cybenko. Iii-conditioning in neural network training problems. *SIAM J. Sci. Stat. Comp.*, 14 :693–714, 1993.
- [115] G. Monari. *Sélection de modèles non-linéaires par leave one out ; étude théorique et application des réseaux de neurones au procédé de soudage par points*. PhD thesis, Université Pierre et Marie Curie, Paris, 1999.
- [116] P. L. Barlett. For valid generalization, the size of the weights is more important than the size of the network. *Neural Information Processing System*, 9, 1997.
- [117] D. M. Allen. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16 :125–127, 1974.
- [118] G. Saporta. *Probabilités, analyse des données et statistique*. 1990.
- [119] D. J. C. McKay. A practical bayesian framework for backpropagation networks. *Neur. Comp.*, 4 :448–472, 1992.
- [120] S. Wold. Personal memories of the early pls development. *Chem. and Int. Lab. Syst.*, 58 :83–84, 2001.
- [121] F. Badran and S. Thiria. Les perceptrons multicouches : de la régression non-linéaire aux problèmes inverses. *J. Phys. IV*, 92 :157–188, 2002.
- [122] Y. Bennani. La sélection de variables. *Systèmes d'apprentissage connexionnistes : sélection de variables, Numéro spécial de la Revue d'Intelligence Artificielle*, 15 :351–371, 2001.
- [123] Y. Le Cun. *Generalization and network design strategies*. Elsevier, 1989.
- [124] Y. Le Cun, J. S. Denker, and S. A. Solla. *Advances in Neural Information Processing Systems 2*. San Mateo, CA, Morgan Kaufman, 1990.
- [125] B. Hassibi and D. Stork. Second order derivatives for networks pruning : Optimal brain surgeon. *Advances in Neural Information Processing Systems*, 4 :950–957, 1992.
- [126] V. Tresp, R. Neuneier, and G. Zimmermann. Early brain damage. *Neur. Inform. Proc. Syst.*, 9 :669–675, 1997.

- [127] M. Yacoub and Y. Bennani. Hvs : A heuristic for variable selection in multilayer artificial neural network classifier. In *IEEE International Conference on Artificial Neural Networks*, 1997.
- [128] M. Yacoub and Y. Bennani. Architecture optimization in feedforward connectionist models. *IEEE Int. Conf. Art. Neur. Net.*, pages 881–886, 1998.
- [129] Y. Bennani. *Apprentissage connexionniste*. Lavoisier, 2006.
- [130] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33 :1065–1076, 1962.
- [131] H. Stoppiglia. *Méthodes statistiques de sélection de modèles neuronaux ; applications financières et bancaires*. PhD thesis, 'Université Pierre et Marie Curie - Paris VI, 1997.
- [132] R. E. Duda and P. E. Hart. *Pattern classification and scene analysis*. John Wiley, 1973.
- [133] J. S. Bridle. *Neurocomputing : Algorithms, Architectures, and Applications*. Springer Verlag, New York, 1990.
- [134] G. E. Moore. Cramming more components onto integrated circuits. *Electr. Mag.*, 19 :114–117, 1965.
- [135] L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll. Generation of periodic surface corrugations. *App. Opt.*, 17, 1978.
- [136] J. B. Lin. Optical lithography - present and future challenges. *C. R. Physique*, 7 :p 858–874, 2006.
- [137] S.-M. Kim, S.-Y. Koo, and C.-M. Lim. Double patterning lithography for dram. *Solid State Technol.*, 2007.
- [138] C. Gwyn, R. Stulen, D. Sweeney, and D. Attwood. Extreme ultraviolet lithography. *J. Vac. Sci. Technol. B*, 16 :3142, 1998.
- [139] J. P. Beasley and D. G. Squire. An electron beam maskmaker. *IEEE Trans. on Elec. Dev.*, 22 :409, 1975.
- [140] L. Pain, S. Tedesco, and C. Constancias. Direct write lithography : the global solution for r&d and manufacturing. *C. R. Physique*, 7 :910–923, 2006.
- [141] S.Y. Chou, P.R. Krauss, and P.J. Renstrom. Imprint of sub-25nm vias and trenches in polymers. *Appl. Phys. Lett.*, 67 :3114, 1995.
- [142] P. Voisin. *Lithographie de nouvelle génération par nanoimpression assistée par UV : étude et développement de matériaux et procédés pour l'application microélectronique*. PhD thesis, Université Joseph Fourier, Grenoble, 2007.
- [143] M. D. Lagerquist, W. Bither, and R. Brouillette. Improving sem linewidth metrology by two-dimensional scanning force microscopy. *Proc. SPIE*, 2725 :494–503, 1996.

- [144] Jérôme Thiault. *Etude par microscopie à force atomique en trois dimensions de l'évolution de la rugosité de bord de ligne lors de la fabrication d'une grille de transistor MOS*. PhD thesis, Université Joseph Fourier, Grenoble, 2007.
- [145] K. P. Bishop, S. M. Gaspar, L. M. Milner, S. S. H. Naqvi, and J. R. McNeil. Grating line shape characterization using scatterometry. In *International Conference on the Application and Theory of Periodic Structures*, J. M. Lerner and W. R. McKinney, eds., *Proc. SPIE*, volume 1545, pages 64–73, 1991.
- [146] S. S. H. Naqvi, J. R. McNeil, R. H. Krukar, and Z. R. Hatab. Grating parameter estimation using scatterometry. In *Miniature and Micro-Optics and Micromechanics*, N. C. Gallagher and C. Roychoudhuri, eds., *Proc. SPIE*, volume 1992, page 170–180, 1992.
- [147] M. R. Murnane, C. J. Raymond, S. L. Prins, S. S. H. Naqvi, and J. R. McNeill. Scatterometry for 0.24 μm –0.70 μm developed photoresist metrology. In *Integrated Circuit Metrology, Inspection, and Process Control IX*, M. H. Bennett, ed., *Proc. SPIE*, volume 2439, page 427–436, 1995.
- [148] B. K. Minhas, S. L. Prins, S. S. H. Naqvi, and J. R. McNeil. Toward sub-0.1-mm cd measurements using scatterometry. In *Metrology, Inspection, and Process Control for Microlithography X*, S. K. Jones, ed., *Proc. SPIE*, volume 2725, pages 729–739, 1996.
- [149] http://www.itrs.net/links/2007itrs/2007_chapters/2007_metrology.pdf.
- [150] B. J. Rice, H. Cao, M. Grumski, and J. Roberts. The limits of cd metrology. *Microelec. Eng.*, 83 :pp 1023–1029, 2006.
- [151] S. Robert, A. Mure Ravaud, S. Reynaud, S. Fourment, F. Carcenac, and P. Arguel. Experimental characterization of subwavelength diffraction gratings by an inverse-scattering neural method. *J. Opt. Soc. Am. A*, 19 :2394–2402, 2002.
- [152] W. Yang, J. Hu, R. Lowe-Webb, R. Korlahalli, D. Shivaprasad, H. Sasano, W. Liu, and D. S. L. Miu. Line-profile and critical-dimension monitoring using a normal incidence optical cd metrology. *IEEE Transactions on Semiconductor Manufacturing*, 17 :pp 564–572, 2004.
- [153] A. Hettwer, N. Benesch, C. Schneider, L. Pfitzner, and H. Ryssel. Phi-scatterometry for integrated linewidth and process control in dram manufacturing. *IEEE Trans. on Sem. Manufact.*, 15 :470–477, 2002.
- [154] H. T. Huang and F. L. Terry. Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and in situ, real-time process monitoring. *Thin Solid Films*, 455-456 :pp 828–836, 2004.
- [155] B. K. Minhas, S. A. Coulombe, S. S. H. Naqvi, and J. R. McNeil. Ellipsometric scatterometry for the metrology of sub 0.1- μm -linewidth structures. *App. Opt.*, 37 :pp 5112–5115, 1998.

- [156] T. Novikova, A. De Martino, S. Ben Hatit, and B. Dré villon. Application of mueller polarimetry in conical diffraction for critical dimension measurements in microelec-tronics. *App. Opt.*, 45 :pp 3688–3697, 2006.
- [157] P. Boher, M. Luet, T. Leroux, J. Petit, P. Barritault, J. Hazart, and P. Chaton. Innovative rapid photo-goniometry method for cd metrology. *Proc. of SPIE Metro-logy, Inspection, and Process Control for Microlithography XVIII*, 5375 :1302–1312, 2004.
- [158] J. Hazart, P. Barritault, S. Garcia, T. Leroux, P. Boher, and K. Tsujino. Robust sub-50-nm cd control by a fast-goniometric scatterometry technique. *Proc. of SPIE Metrology, inspection, and process control for microlithography XXI*, 6518, 2007.
- [159] X. Niu and Y. Sanjay. Specular spectroscopic scatterometry in duv lithography. *Proceedings of SPIE Metrology, Inspection, and Process Control for Microlithography XIII*, 3677 :159–168, 1999.
- [160] S. S. H. Naqvi, S. Kaspar, K. Hickman, K. Bishop, and J. R. McNeil. Linewidth measurement of gratings on photomasks : a simple technique. *App. Opt.*, 31 :1377–1384, 1992.
- [161] S. Soulan, M. Besacier, P. Schiavone, and T. Leveder. Real-time profile shape reconstruction using dynamic scatterometry. *Proc. SPIE Advanced Lithography*, 2007.
- [162] D. Fuard, C. Perret, V. Farys, C. Gourgon, and P. Schiavone. Measurement of residual thickness using scatterometry. *J. Vac. Sc. Tech. B*, 23 :3069–3074, 2005.
- [163] E. Liu and F. L. Terry. Immersion scatterometry for improved nano-scale topogra-phy measurements. *Phys. Stat. Sol. a*, 205 :784–788, 2008.
- [164] M. Foldyna, A. De Martino, E. Garcia-Caurel, R. Ossikovski, C. Licitra, F. Bertin, K. Postava, and B. Drevillon. Critical dimension of biperiodic gratings determined by spectral ellipsometry and mueller matrix polarimetry. *Eur. Phys. J. App. Phys.*, 42 :351–359, 2008.
- [165] I. Kallioniemi, J. Saarinen, and E. Oja. Optical scatterometry of subwavelength diffraction gratings : neural-network approach. *App. Opt.*, 37 :pp 5830–5834, 1998.
- [166] I. Kallioniemi, J. Saarinen, and E. Oja. Characterization of diffraction gratings in a rigorous domain with optical scatterometry. *App. Opt.*, 38 :pp 5920–5930, 1999.
- [167] S. Robert, A. Mure-Ravaud, and D. Lacour. Characterization of optical diffraction grating by use of a neural method. *J. Opt. Soc. Am. A*, 19 :pp 24–32, 2002.
- [168] R. Quintanilha, P. Thony, D. Henry, and J. Hazart. 3d-features analysis using spectroscopic ellipsometry. *Proc. of SPIE Metrology, Inspection, and Process Control for Microlithography XVII*, 5375 :456–467, 2004.
- [169] R. Quintanilha, J. Hazart, P. Thony, and D. Henry. Application of spectroscopic scatterometry method in hole matrices analysis. *Proc. of SPIE Metrology, Inspec-tion, and Process Control for Microlithography XIX*, 5752 :204216, 2005.

- [170] B. Yaakovovitz, Y. Cohen, and Y. Tsur. Line edge roughness detection using deep uv light scatterometry. *Microelcectr. Eng.*, 84 :619–625, 2007.
- [171] P. Boher, J. Petit, T. Leroux, J. Foucher, Y. Desières, J. Hazart, and P. Chaton. Optical fourier transform scatterometry for ler and lwr metrology. *Proc. of SPIE Metrology, Inspection, and Process Control for Microlithography XIX*, 5752 :192–203, 2005.
- [172] L. Mashev and S. Tonchev. Formation of holographic diffraction gratings in photoresist. *App. Phys. A*, 26 :143–149, 1981.
- [173] F. Lazzarino, C. Gourgon, Schiavone, and C. Perret. Mold deformation in nanoimprint lithography. *J. Vac. Sci. Technol. B*, 22 :3318–32322, 2004.
- [174] S. Robert, A. Mure-Ravaud, S. Thyria, M. Yacoub, and F. Badran. Neural selection of the optimal signature for a rapid characterization of a submicrometer period grating. *Opt. Comm.*, 238 :pp 215–228, 2004.
- [175] S-C. Wang, Y-S. Ku, C-H. Ko, and N. Smith. Method of characteristic signatures matching through discrepancy enhanced library generation process. *Proc. SPIE*, 5908 :1–11, 2005.
- [176] Y. S. Ku, S. C. Wang, D. M. Shyu, and N. Smith. Scatterometry-based metrology with featureregion signature matching. *Opt. Exp.*, 14 :pp 8482–8491, 2006.
- [177] E. M. Drège, J. A. Reed, and D. Byrne. Linearized information of scatterometric data to obtain surface profile information. *Opt. Eng.*, 41 :pp 225–236, 2002.
- [178] C-H. Ko, Y-S. Ku, and N. Smith. Correlation os scatterometry sensitivities to variation in device parameters. *Proc. SPIE*, 6152, 2006.
- [179] N. Destouches, C-A. Guerin, M. Lequime, and H. Giovannini. Determination of the phase of the diffracted field in the optical domain : Application to the reconstruction of surface profiles. *Opt. Comm.*, 198 :233–239, 2001.
- [180] S. Robert, Y. Battie, D. Jamon, and F. Royer. Accurate and rapid optical characterization of an anisotropic guided structure based on a neural network. *App. Opt.*, 46 :2036–2040, 2007.

Résumé

La miniaturisation est actuellement la voie de recherche la plus explorée dans divers domaines de la science et de la technologie. Les processus de fabrication comme la lithographie se sont prodigieusement développés au cours de ces dernières années et permettent ainsi une réduction importante de la taille des composants. Ce progrès technologique a créé systématiquement le besoin de techniques de caractérisation fiables, efficaces et si possible à moindre coût.

L'objet de cette thèse porte sur l'étude d'un outil mathématique original, à savoir les réseaux de neurones, dans le cadre de la métrologie optique et dimensionnelle achevée par voie ellipsométrique. Dans le premier volet de ce travail nous avons montré que le réseau de neurones peut être efficacement employé pour la détermination des propriétés optiques et géométriques (indice de réfraction et épaisseur) des couches minces. A titre illustratif, la classification neuronale a été proposée pour estimer la gamme d'épaisseur des couches sans aucune information a priori sur la structure. Cette technique peut être couplée avec n'importe quel autre algorithme d'optimisation nécessitant une connaissance préalable de la solution. Le second volet montre clairement l'apport des réseaux de neurones dans le domaine de la scatterométrie pour la caractérisation des réseaux de diffraction possédant différents profils géométriques. La méthode neuronale peut également être employée pour la détermination de la période du réseau lorsque cela est nécessaire. La classification neuronale a ensuite été appliquée pour l'identification structurale du modèle géométrique, donnant ainsi une application directe en lithographie pour la détection automatique d'une couche résiduelle nuisible à l'étape de gravure.

Mots-clés : Ellipsométrie, réseaux de neurones, scatterométrie, problème inverse, structures périodiques diffractantes, couches minces, métrologie optique et dimensionnelle.

Abstract

Nowadays, miniaturization is the most explored research topic in various domains of science and technology. Manufacturing processes, such as lithography, have been prodigiously developed during these last years and allow high scale integration of devices. This technological progress has created systematically the need of reliable, efficient and if possible low cost characterization techniques.

The aim of this PhD is to study and implement an original mathematical tool, namely neural networks, within the framework of optical and dimensional metrology achieved by ellipsometry. In the first part of this work, we have shown that the neural network can be effectively employed for the determination of the optical and geometrical properties (refractive index and thickness) of thin films. For instance, neural classification has been proposed in order to estimate the thickness range of films without any prior information about the structure. This technique can be coupled with any other optimization algorithm requiring a prior knowledge about the solution. In the second part, we have clearly shown the contribution of the neural network in scatterometry for the characterization of diffraction gratings with different geometrical profiles. The neural method can also be employed to determine the grating pitch when it is required. Neural classification has been applied for structural identification of the geometrical model, giving thus a direct application in lithography for automatic detection of the residual layer undesirable for the etching step.

Keywords : Ellipsometry, neural networks, scatterometry, inverse problem, periodic diffractive structures, thin films, optical and dimensional metrology.